Clinical Infectious Diseases
15 December 2012
Volume 55
Supplement 4
Volume 55 Supplement 4 15 December 2012 Pages S215–S335
IDSA Infectious Diseases Society of America
hivma
OXFORD UNIVERSITY PRESS
cid.oxfordjournals.org
OXFORD

Clinical Infectious Diseases

The Global Enteric Multicenter Study (GEMS)

A Supplement to *Clinical Infectious Diseases*

Clinical Infectious Diseases

The Global Enteric Multicenter Study (GEMS)

Senior Supplement Editor: Myron M. Levine, MD, DTPH

Associate Supplement Editors: Karen L. Kotloff, MD
James P. Nataro, MD, PhD, MBA
Roy M. Robins-Browne, MBBCH, PhD
Claudio F. Lanata, MD, MPH

This supplement was sponsored by the Bill & Melinda Gates Foundation.

Clinical Infectious Diseases

15 December 2012
Volume 55
Supplement 4

The Global Enteric Multicenter Study (GEMS)

SUPPLEMENT ARTICLE

The Global Enteric Multicenter Study (GEMS): Impetus, Rationale, and Genesis

Myron M. Levine,[1] Karen L. Kotloff,[1] James P. Nataro,[2] and Khitam Muhsen[1]

[1]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; and [2]Department of Pediatrics, University of Virginia School of Medicine, Charlottesville

Diarrheal disease remains one of the top 2 causes of young child mortality in the developing world. Whereas improvements in water/sanitation infrastructure and hygiene can diminish transmission of enteric pathogens, vaccines can also hasten the decline of diarrheal disease morbidity and mortality. From 1980 through approximately 2004, various case/control and small cohort studies were undertaken to address the etiology of pediatric diarrhea in developing countries. Many studies had methodological limitations and came to divergent conclusions, making it difficult to prioritize the relative importance of different pathogens. Consequently, in the first years of the millennium there was no consensus on what diarrheal disease vaccines should be developed or implemented; however, there was consensus on the need for a well-designed study to obtain information on the etiology and burden of more severe forms of diarrheal disease to guide global investment and implementation decisions. Accordingly, the Global Enteric Multicenter Study (GEMS) was designed to overcome drawbacks of earlier studies and determine the etiology and population-based burden of pediatric diarrheal disease. GEMS, which includes one of the largest case/control studies of an infectious disease syndrome ever undertaken (target approximately 12 600 analyzable cases and 12 600 controls), was rolled out in 4 sites in sub-Saharan Africa (Gambia, Kenya, Mali, Mozambique) and 3 in South Asia (Bangladesh, India, Pakistan), with each site linked to a population under demographic surveillance (total approximately 467 000 child years of observation among children <5 years of age). GEMS data will guide investment and help prioritize strategies to mitigate the morbidity and mortality of pediatric diarrheal disease.

In the 55 years between the end of World War II and the close of the 20th century, developing countries, including many newly established nations that emerged from the dissolution of colonial empires, grappled with growing their economies and improving the health of their people. While progressive economic development ensued in many countries (and was impressive in some), others countries notably lagged. By the late 1990s, the United Nations (UN) categorized a subset of approximately 43–50 as the "least developed countries," many of which were located in sub-Saharan Africa and some parts of Asia [1]. These least developed countries, in particular, were characterized by extremely low gross national income per capita, high young child (<age 5 years) mortality, low adult (particularly female) literacy, and abbreviated adult life expectancy [2, 3]. Diarrheal diseases, pneumonia, measles, and malaria were typically among the top causes of young child mortality. In general, the higher the infant and young child mortality rate, the larger the fraction of mortality attributed to diarrheal diseases. Estimates of global young child (<age 5 years) mortality suggest that in the early years of the millennium an estimated 10.6 million young child deaths occurred annually [4, 5], with approximately 17%–21% of deaths due to diarrheal disease [4, 6, 7] and approximately 70% of all diarrheal mortality localized in 15 countries in Africa and South and Southeast Asia. Addressing the main causes of young child mortality

Correspondence: Myron M. Levine, MD, DTPH, Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201 (mlevine@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S215–24

DOI: 10.1093/cid/cis761

in developing countries, including diarrheal diseases, became a global priority [8].

MOBILIZATION IN THE NEW MILLENNIUM

Circa 2000, 3 new entities came on the scene that rapidly interrelated in a synergistic way to offer extraordinary potential to accelerate the decline of young child mortality in developing countries, and particularly the component due to diarrheal diseases. In 2000, the 55th Session of the UN General Assembly adopted the UN Millennium Declaration [9], committing the countries of the world to mobilize resources to reduce poverty and improve health and education by 2015, with progress judged by whether or not certain specific goals were attained. One of these, Millennium Development Goal #4, aims to reduce young child mortality by 67% by 2015, compared to the 1990 baseline.

Second, in 1999 the nascent Bill & Melinda Gates Foundation entered the arena of global health and brought zeal, commitment, and passionate advocacy, as well as substantial new financial resources, to improve the survival of young children in developing countries. Finally, at the World Economic Forum in February 2000, the Global Alliance for Vaccines and Immunization (GAVI, now called the GAVI Alliance) was launched. The GAVI Alliance is a consortium that consists of UN agencies (World Health Organization [WHO], United Nations Children's Fund [UNICEF], World Bank) involved with immunization, vaccine supply, and vaccine financing; developing and donor countries; the vaccine industry (in both industrialized and developing countries); technical and research institutes; civil society; and the Bill & Melinda Gates Foundation and other private philanthropic foundations. In its decade of existence, GAVI has been highly successful in strengthening the delivery of immunization services and in introducing life-saving new vaccines into the Expanded Programme on Immunization of many of the poorest countries of the world, including those in sub-Saharan Africa. In 2002, GAVI established and funded 2 Accelerated Development and Introduction Plans (ADIPs), one for rotavirus vaccine and the other for pneumococcal conjugate vaccines. The fundamental aims of the rotavirus ADIP were (1) to provide information (eg, documentation of the safety, immunogenicity, and efficacy of rotavirus vaccines in infants in developing countries) that enables evidence-based decisions regarding the use of rotavirus vaccines, and (2) to accelerate the availability of new rotavirus vaccines appropriate for use in developing countries.

GEOGRAPHIC FOCUS

In order to intervene in a strategic way to accelerate the decline of young child mortality globally, efforts must be concentrated in 2 main geographic areas: sub-Saharan Africa, where 33 of the 35 countries with world's highest young child mortality rates are located [2, 3, 10–12], and South Asia, where the size of the young child population is enormous, leading to

Table 1. The 5 Main Clinical Syndromes of Diarrheal Disease Seen Among Infants and Young Children Presenting to Health Centers and Hospitals in Developing Countries

Clinical Syndrome	Characteristic Signs and Symptoms	Proportion of Pediatric Diarrhea Patients Presenting to Health Facilities Who Exhibit This Syndrome	Some Etiologic Agents Associated With This Syndrome
Simple gastroenteritis	Loose stools (often with mucus but no blood), occasional vomiting, anorexia, low-grade fever, malaise	80%–85%	Rotavirus, ETEC, EPEC, *Cryptosporidium*, norovirus GI and II, enteric adenovirus types 40 & 41
Dysentery	Gross blood in loose stools (that may be quite scanty), fever (sometimes high), abdominal cramps, tenesmus; many dysentery patients appear clinically toxic	5%–15%	*Shigella*, *Campylobacter jejuni*, *Entamoeba histolytica*, nontyphoidal *Salmonella*
Profuse purging	Passage of copious watery stools that may resemble "rice water"; fever, if present, is typically low grade; signs of overt dehydration (diminished skin turgor, sunken eyes, dry mucous membranes) may be present	<5%	*Vibrio cholerae* O1 & O139, ETEC
Persistent diarrhea	The same signs and symptoms as simple gastroenteritis but diarrhea continues unabated for at least 14 days	<5%	*Giardia lamblia*, EPEC
Acute vomiting	Acute onset of nausea and repetitive vomiting with little or no diarrhea	~1%–2%	Norovirus

See Kotloff et al in this supplement for precise clinical definitions used in the GEMS study.

Abbreviations: EPEC, enteropathogenic *Escherichia coli*; ETEC, enterotoxigenic *Escherichia coli*; GEMS, Global Enteric Multicenter Study.

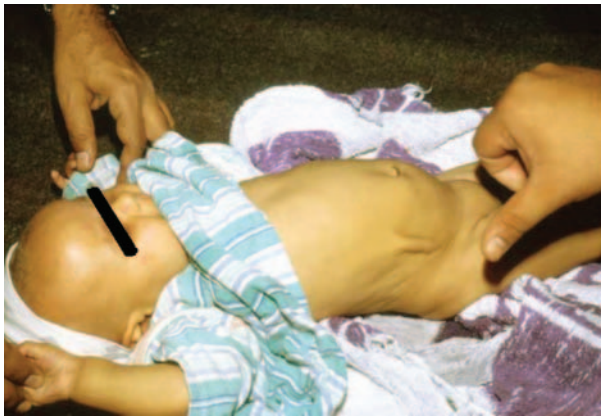

Figure 1. An infant who presented with diarrheal dehydration consequent to simple gastroenteritis that was not treated promptly or effectively. Loss of turgor of skin over the abdomen is visible as "tenting," following pinching. Simple gastroenteritis caused by many etiologic agents in young infants in developing countries can lead to dehydration. The fundamental reason is that per kilogram of body weight, the daily water and electrolyte requirements of young infants are substantially greater than those of older children. Thus, abnormal losses from diarrhea, vomiting, and fever, accompanied by inadequate fluid intake and lack of prompt and appropriate replacement (as with glucose/electrolyte oral rehydration solution), can lead to moderate and severe dehydration and death. This photograph was kindly provided by Dr Dipika Sur of the National Institute of Cholera and Enteric Diseases, Kolkata, India.

Figure 2. Dysentery is diagnosed clinically as the presence of gross blood in diarrheal stools. Dysentery stools can be quite scanty and composed mainly of mucus and blood (shown here). Bacillary dysentery is typically preceded by 18–24 hours of watery diarrhea, accompanied by high fever and toxemia, before the loose stools become scanty and bloody. Dysentery indicates substantial damage to the mucosa of the colon and terminal ileum.

a large number of deaths, despite the mortality rates being lower than in sub-Saharan Africa [2, 3, 10–12].

CLINICAL SYNDROMES OF PEDIATRIC DIARRHEAL DISEASE IN DEVELOPING COUNTRIES

As seen by clinical health providers at fixed healthcare facilities in developing countries, almost all cases of pediatric diarrheal illness can be conveniently characterized as falling into 1 of 5 clinical syndromes [13] (Table 1). Approximately 80%–85% of patient episodes present as "simple gastroenteritis" with the subject passing loose or watery stool (often with mucus but not with blood), low-grade fever, occasional vomiting, anorexia, and apparent malaise (Figure 1). Approximately 5%–15% of children present with overt dysentery (gross blood in the diarrheic stools) (Figure 2), often accompanied by fever (sometimes high); many dysenteric patients appear clinically toxic. A small proportion of cases in older children present with profuse watery diarrhea, passing such voluminous "rice water" stools that even older children can rapidly become severely dehydrated (Figure 3). Another few percent of pediatric cases present with a history of apparent simple gastroenteritis that began 14 or more days previously but did not abate [14]; this defines "persistent diarrhea," a syndrome that particularly can have adverse nutritional consequences [15]. Finally, a few percent of children are brought by caretakers for care because of vomiting rather than diarrhea as the main complaint. Few reports have described expanded etiologic analyses in relation to these clinical syndromes.

THE INTERRELATIONSHIP BETWEEN NUTRITIONAL STATE AND DIARRHEAL DISEASE

It has long been recognized that there is an intimate relationship between diarrheal disease and undernutrition in pediatric populations in developing countries [16, 17]. Diarrheal disease, with its injury to the gut, can lead children to fall off their growth curve. Conversely, more extreme forms of chronic malnutrition predispose young children to diarrhea-related mortality. For example, moderate and severe stunting is a strong risk factor for death from diarrheal disease [18].

LESSONS FROM THE EARLY 20TH CENTURY IN NORTH AMERICA AND EUROPE

Mortality from diarrheal disease is currently extremely low in industrialized countries, but it was a vexing public health problem a century ago when populations in current industrialized countries lived in conditions resembling those endured by people in developing countries today [19–23]. In fact, wherever populations live in crowded conditions marked by

Figure 3. A Bangladeshi child with cholera is shown who experienced copious purging of rice water stools prior to presenting with severe dehydration. The child, with deeply sunken eyes, is lying on a cholera cot with his watery stools being collected in a bucket for measurement of volume (to guide replacement therapy). After rapid replacement of the child's fluid and electrolyte deficits with intravenous fluids, the health worker is attempting to transition the child to oral rehydration fluids administered by his caretaker, under supervision.

widespread fecal contamination, lack of treated water supplies and sanitation to remove human fecal waste, and lack of refrigeration to preserve food, the transmission of bacterial, viral, and protozoal enteric pathogens is enhanced and pediatric diarrheal disease can rage rampant. A shared vision of the Millennium Declaration is that all countries will undergo accelerated development such that with improved housing, provision of sanitation and safe water, enhanced food safety, and access to primary health care, diarrheal disease and pneumonia mortality will plummet. While that is the ultimate aim, it may be possible to accelerate markedly the decline in diarrheal disease mortality by certain cross-cutting general interventions (such as improved treatment of diarrhea and focused water/sanitation/hygiene improvements) and by immunizing infants and young children against the major etiologic agents responsible for clinically severe and fatal forms of diarrheal disease.

In the early years of the millennium, other than vaccines against rotavirus, there was not a broad consensus on what other diarrheal disease vaccines should be high priority for development and accelerated introduction, given the limited resources and supply issues pertinent at the global level. One must also recognize that for rotavirus vaccines there were mature industrialized country markets waiting to reward companies that invested in rotavirus vaccines and achieved licensure for their products in North America, Europe, and Australia. This guaranteed the development of these vaccines, a situation not operative for pathogens prevalent in developing countries but uncommon in industrialized countries (eg, *Shigella* species, enterotoxigenic *Escherichia coli* [ETEC]).

WHY KNOWLEDGE OF THE SPECIFIC ETIOLOGY OF PEDIATRIC DIARRHEAL DISEASE IN DEVELOPING COUNTRIES IS IMPORTANT

In developing country pediatric populations, it has long been recognized that there is a striking association between measles, diarrhea, and mortality [24, 25], and measles vaccine in such populations has been referred to as the first diarrheal disease vaccine [26]. The impressive and precipitous decline of measles as a cause of global young child mortality consequent to repetitive mass immunization campaigns with measles vaccine [27], particularly in sub-Saharan Africa, has led many to hypothesize that a sizable reduction of diarrheal disease mortality might be achievable if the specific major offending diarrheal pathogens were clearly elucidated and if vaccines existed and could be delivered to populations at high risk. And if licensed vaccines against those pathogens did not exist, advocacy could be undertaken to accelerate or initiate their development. Regrettably, as discussed below, as of the first years of the millennium, these data were not available with the precision necessary to drive investment decisions and to establish implementation priorities. In the 1980s and 1990s, in the absence of a robust evidence base, a Steering Committee on Diarrheal Diseases of WHO, following Delphian deliberations, proposed that the highest priority vaccines needed to prevent diarrheal diseases in developing countries were ones against rotavirus, ETEC, *Shigella* species, and *Vibrio cholerae* O1.

EARLY STUDIES INVESTIGATING THE ETIOLOGY OF SEVERE AND FATAL FORMS OF DIARRHEAL DISEASE IN YOUNG CHILDREN IN DEVELOPING COUNTRIES

Studies attempting to define the etiology of pediatric diarrheal disease in developing countries have been carried out for

many decades. In the 1950s and early 1960s these studies were hampered by the fact that only a relatively few diarrheal pathogens were recognized and they were recovered from only a small proportion of diarrhea cases [28–33]. Thus, in that period the urgent need was to identify the etiologic agents. The 1970s and 1980s ushered in an age in which a plethora of new enteric pathogens were described including ETEC, rotavirus, *Campylobacter jejuni*, enteric adenovirus serotypes 40 and 41, what came to be known as noroviruses, astrovirus, enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), enterohemorrhagic *E. coli* (EHEC), diffuse adherence *E. coli*, and *Cryptosporidium* species, to name some. In early studies, some of these agents were detected in a proportion of cases of pediatric diarrhea in developing countries.

For some years practical, robust, economical tests to detect even relatively common etiologic agents, such as ETEC, enteropathogenic *E. coli* (EPEC), rotavirus, EAEC, and norovirus, remained unavailable. Thus, for some agents, animal models [34], electron microscopy [35], laborious fecal concentration followed by acid fast, Giemsa, or fluorescent staining and direct examination by a skilled light microscopist [36–38], or complicated competitive enzyme-linked immunosorbent assays [39] were required, making large-scale comparisons impractical. However, with time, improved (often commercial) diagnostics became available to detect some of these pathogens with a high degree of standardization, thus enabling comparisons of etiology across geographic sites and over time. In particular, the advent of nucleic acid-based testing revolutionized the landscape, initially with DNA hybridization probes [40–43], then with iterations of polymerase chain reaction (PCR) (including multiplex techniques) [44–46; Panchalingam et al, this supplement] and quantitative reverse transcriptase PCR (to detect RNA viruses). Advances were also made in diagnostic methods to detect protozoal enteropathogens such as *Cryptosporidium* species and *Entamoeba histolytica*, including highly standardized, practical commercial immunoassay kits [47, 48].

MODERN STUDIES OF THE ETIOLOGY OF DIARRHEA IN YOUNG CHILDREN IN DEVELOPING COUNTRIES

By reviewing studies carried out since 1980, one identifies a number that employed tests for many of the "modern" etiologic agents. One might assume, therefore, that one can derive a clear landscape of the major enteric pathogens responsible for diarrheal disease of a severity that might lead to death in the geographic areas of highest mortality risk for young children. In fact, while there are indeed reports, most have notable shortcomings that limit their utility to address the question at a global level. For example, while there have been many studies of the etiology of pediatric diarrhea, relatively few have been performed in sites with very high or high young child mortality [49–78], as defined by UNICEF [79]. In particular, very few studies were carried out in sub-Saharan Africa [51–54, 57–59, 64–67]. Although a number of studies enrolled subjects at several sites within a single country, a multinational study such as that sponsored by WHO and reported by Huilan et al was a rare exception [55]. Most studies examining the etiology of pediatric diarrhea limited enrollment to infants and toddlers <24 months of age [61, 66, 74, 75, 80–84] or occasionally to children up to 35 months of age [49, 55, 56, 85]. Few studies enrolled children through 59 months of age, which could capture pathogens such as *V. cholerae* O1 or O139, which are more heavily represented in older preschool children with severe diarrhea (who comprise a potential target group for prevention).

Because the transmission of many diarrheal pathogens is highly seasonal and since there may also be considerable year-to-year variation in the relative frequency with which they circulate, it is important that studies of the etiology of pediatric diarrhea take this into account and be performed over a period of at least 2 and preferably 3 years. Some studies enrolled for <6 calendar months [57, 65, 66, 69, 82, 86], others for 6–24 months [50–56, 58–60, 62, 63, 67, 68, 71–75, 80, 81, 84, 87–97]. A few studies proceeded for 24–36 months [49, 81, 83, 85, 98–100] and 3 studies enrolled for >36 calendar months [70, 74, 97].

Approximately one-half of the studies investigating the etiology of pediatric diarrhea in developing countries mentioned above also sought pathogens in matched or relevant control subjects [49, 50, 52, 55, 57–60, 62, 63, 65, 68, 69, 73, 82, 84, 85, 87–97, 100, 101]. Some case/control studies were linked to a large defined population that had undergone a detailed recent census or that was under prospective demographic surveillance so that population-wide incidence rates could potentially be calculated; some cohort studies were also nested within such defined populations to allow potential extrapolations of incidence to the larger population. However, no case/control study recorded a baseline survey to estimate the healthcare seeking patterns and preferences of the larger population served by the hospitals, health centers, or other sites where enrollment of patients was carried out.

Most studies looked for an array of enteric pathogens that by that time were widely regarded as being associated with pediatric diarrhea in developing countries, such as rotavirus, ETEC, EPEC, EAEC, *Shigella* species, nontyphoidal *Salmonella*, *C. jejuni*, *V. cholerae* (usually in Asian studies), *Cryptosporidium* species, and *Giardia* species. In addition, some tested for 1 or more of the following: EIEC, diffusely adherent *E. coli*, EHEC, *Aeromonas hydrophila*, *Plesiomonas shigelloides*, enterotoxigenic *Bacteroides fragilis*, *Clostridium difficile* toxin,

noroviruses, enteric adenoviruses, *E. histolytica*, *Cyclospora cayetanensis*, *Strongyloides stercoralis*. Many studies characterized ETEC isolates by toxin types, that is, those that elaborate heat-stable or heat-labile enterotoxin only, or those that produce both; a proportion of studies serogrouped *Shigella* isolates. However, few reports characterized ETEC by the fimbrial colonization factors that they express or fully serotyped *Shigella* isolates. Such information is important to guide vaccine development.

Among the post-1980 case/control reports of the broad etiology of pediatric diarrhea in developing countries, none related etiology to the different clinical syndromes of diarrheal disease and none described a follow-up visit (or visits) after a period of 1–2 months to ascertain whether the child was still alive and whether overt sequelae were evident. Few studies enrolled enough subjects to assure reasonable statistical power to detect significant differences in the rate of isolation in cases versus controls and to allow the calculation of odds ratios to assess the degree of pathogenicity by the strength of association.

THE GENESIS OF THE GEMS

Despite the many publications on the etiology of pediatric diarrheal disease, the recognition in the first years of the millennium of the existence of so many different potential diarrheal pathogens, the limitations of most of the studies and the great variation in results and conclusions made it impossible to set priorities on what enteric vaccines or other specific interventions were most needed to control morbidity and mortality in developing countries. A consensus emerged in the enteric disease research and disease control communities on the need for a definitive multicenter study that would attempt to address all the limitations of previous studies. An exhortation was made to design, organize, and undertake a large, well-powered, case/control study of the etiology and burden of pediatric diarrheal disease in multiple sites of high mortality, particularly in sub-Saharan Africa and South Asia [13]. It was proposed that the study should use state-of-the-art microbiological methods to detect a wide array of pathogens in patients whose clinical syndromes of presentation were carefully documented [13] and to perform the study in a defined population. It was also urged that a novel design be utilized that included a follow-up visit to case and control households 1–2 months after enrollment to ascertain whether there was mortality that occurred beyond the peri-enrollment period [13].

In 2004 the Bill & Melinda Gates Foundation made a strategic decision to expand its portfolio of projects in the area of enteric diseases, recognizing that these illnesses were one of the major killers of young children. At the behest of the Foundation, the Center for Vaccine Development of the University of Maryland School of Medicine submitted a proposal to undertake a definitive, multicenter, 3-year, highly-powered case/control study to determine the diarrheal pathogens that exact the highest burden of morbidity, mortality, and nutritional consequences in 3 different pediatric age strata (0–11, 12–23, and 24–59 months) in multiple sites in sub-Saharan Africa and South Asia (Table 2), with each site linked to a defined population under ongoing demographic surveillance (total of approximately 467 000 child-years of observation over 36 calendar months among the 7 sites) so that population-based incidence rates could be calculated, and to link etiology to clinical syndrome. The project, which was funded in 2006, would utilize optimal clinical and laboratory methods standardized across the different study sites. Officially designated

Table 2. Several Salient Features of the 7 Field Sites of the Global Enteric Multicenter Study

Country	Collaborating Institution	Field Site	Setting	Annual Young Child (<5 y) Population Under Demographic Surveillance[a]
The Gambia	Medical Research Council Unit, The Gambia	Basse (Upper River Division)	Rural	29 076
Kenya	CDC/Kenya Medical Research Institute (KEMRI) Research Station	Nyanza Province	Rural	21 603
Mali	Centre pour le Développment des Vaccins du Mali (CVD-Mali)	Djikoroni Para and Banconi quartiers, Bamako	Urban	31 768
Mozambique	Centro de Investigação em Saúde de Manhiça (CISM)	Manhiça District	Rural	15 380
Bangladeshm	International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B)	Mirzapur Sub-District, Tangail District	Rural	25 560
India	National Institute of Cholera and Enteric Diseases (NICED)	Wards 14, 31, 34, 58, & 59	Urban	13 416
Pakistan	Aga Khan University	Coastal settlements 20 km south of Karachi	Periurban	25 659

[a] Median No. of children <5 years of age in the population at each GEMS site based on multiple rounds of demographic surveillance.

"Diarrheal Disease in Infants and Young Children in Developing Countries," the project came to be known as the Global Enteric Multicenter Study (GEMS). The keystone component of GEMS is one of the largest case/control studies ever carried out of an infectious disease syndrome, with a target enrollment of 600 analyzable cases of moderate-to-severe diarrhea (defined by Kotloff et al in this supplement) per each of 3 age strata (0–11, 12–23, and 24–59 months), per each of 7 sites, over 3 years (total of approximately 12 600 analyzable cases) and a similar number of matched controls. Additional subaims of the GEMS include the identification of water/sanitation/hygiene risk factors for specific pathogens, quantification of the economic burden of pediatric diarrhea on poor households in sub-Saharan Africa and Asia, full serotyping of *Shigella* isolates, elucidation of the fimbrial colonization factor antigen types of ETEC strains, and genotypic or further characterization of other major enteropathogens identified. This initial 3-year case/control study of moderate-to-severe diarrhea has been coined GEMS-1. A subsequent 1-year follow-on study in the same 7 sites that is investigating the etiology of less severe diarrhea, as well as moderate-to-severe diarrhea, is referred to as GEMS-1A.

In this supplement, contributing papers describe basic assumptions that guided the GEMS-1 study design (Farag et al); the selection of the 7 GEMS-1 sites and the clinical and epidemiologic methods (Kotloff et al); the biostatistical strategies to analyze the data (Blackwelder et al); the data management methods needed to handle the enormous quantities of data (Biswas et al); and an innovative approach that uses the cohorts of cases and controls prospectively followed for approximately 60 days after enrollment into GEMS-1 and weighted generalized linear model regression to estimate the association between exposures recorded during the case/control component and outcomes detected during the follow-up (Sommerfelt et al). Additional papers provide a detailed review of the published literature accompanied by meta-analyses to examine the association between *Giardia lamblia* and acute and persistent diarrhea (Muhsen and Levine); the standardized laboratory methods used to identify diarrheal pathogens (Panchalingam et al); factors that explain the excretion of enteric pathogens by persons without diarrhea (Levine and Robins-Browne); laboratory diagnostic challenges in case/control studies of diarrhea in developing countries (Robins-Browne and Levine); and analyses of the economic burden of diarrheal disease at 6 of the 7 GEMS sites in Africa and Asia just prior to initiation of the case/control studies (Rheingans et al).

It is anticipated that the GEMS data will help to guide investment and implementation decisions in the area of diarrheal diseases on the global level. The GEMS consortium can also serve as a platform in the future to evaluate various interventions against diarrheal diseases (vaccines, water/sanitation hygiene improvements, novel therapies, diagnostics) at multiple sites, simultaneously. In this way the time required to obtain definitive answers can be diminished.

Notes

Acknowledgments. The authors thank Drs Thomas Brewer, Regina Rabinovich, Jan Agosti, and Niranjan Bose and Abby Dain of the Bill & Melinda Gates Foundation for their steadfast support and encouragement. Thanks are also due to Gloria Smedley, Jean Flanders Armstrong, and Dottie Small of the Center for Vaccine Development for their tireless and invaluable administrative support. Finally, we greatly appreciate the encouragement, sage advice, and collegial friendship of Professors George Griffin (St George's Hospital, London), Fred Binka (University of Ghana, Accra), and Zulfiqar Bhutta (Aga Khan University, Karachi), co-chairs of the GEMS International Strategic Advisory Committee.

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. United Nations Conference on Trade and Development Secretariat. The least developed countries report 2002. Geneva, Switzerland: United Nations, **2002**.
2. United Nations Children's Fund. State of the World's Children 2000. New York, NY: UNICEF, **2000**.
3. United Nations Children's Fund. State of the world's children 2002. Leadership. New York, NY: UNICEF, **2002**.
4. Bryce J, Boschi-Pinto C, Shibuya K, Black RE. WHO estimates of the causes of death in children. Lancet **2005**; 365:1147–52.
5. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet **2010**; 375:1969–87.
6. Parashar UD, Hummelman EG, Bresee JS, Miller MA, Glass RI. Global illness and deaths caused by rotavirus disease in children. Emerg Infect Dis **2003**; 9:565–72.
7. Boschi-Pinto C, Velebit L, Shibuya K. Estimating child mortality due to diarrhoea in developing countries. Bull World Health Organ **2008**; 86:710–7.
8. World Health Organization. Removing obstacles to healthy development. Report on infectious diseases. **1999**. Report No. WHO/CDS/99.1.
9. United Nations Millennium Declaration, A/RES/55/2. United Nations General Assembly, **2000**.
10. United Nations Children's Fund. State of the world's children 2011. Adolescence, an age of opportunity. New Jersey: Hatteras Press, **2011**:1–128.
11. United Nations Children's Fund. State of the world's children 2008. Child survival. New York: Hatteras Press, **2007**.
12. United Nations Children's Fund. State of the world's children 2009. In: Moccia P, ed. Maternal and Newborn Health. Sterling, VA: Colorcraft of Virginia, **2008**.
13. Levine MM. Enteric infections and the vaccines to counter them: future directions. Vaccine **2006**; 24:3865–73.
14. Lanata CF, Black RE, Gilman RH, Lazo F, Del AR. Epidemiologic, clinical, and laboratory characteristics of acute vs persistent diarrhea in periurban Lima, Peru. J Pediatr Gastroenterol Nutr **1991**; 12:82–8.

15. Cruz JR, Pareja G, Caceres P, Cano F, Chew F. Acute and persistent diarrheal disease and its nutritional consequences in Guatemalan infants. Arch Latinoam Nutr **1989**; 39:263–77.
16. Mata LJ. The children of Santa Maria Cauque. Cambridge, MA: MIT Press, **1978**.
17. Scrimshaw NS. Historical concepts of interactions, synergism and antagonism between nutrition and infection. J Nutr **2003**; 133:316S–321S.
18. Black RE, Allen LH, Bhutta ZA, et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet **2008**; 19;371:243–60.
19. Stern A. Infant mortality in summer: how to conduct a successful campaign against it. Arch Pediatr **1911**; 28:287–94.
20. Levine MM, Robins-Browne R. Vaccines, global health and social equity. Immunol Cell Biol **2009**; 87:274–8.
21. Levine MM, Edelman R. Acute diarrheal infections in infants. I. Epidemiology, treatment and immunoprophylaxis. Hospital Practice **1979**; 14:89–100.
22. Levine MM, Losonsky G, Herrington D, et al. Pediatric diarrhea: the challenge of prevention. Pediatric Infectious Disease **1986**; 5: S29–43.
23. Meyer EC. Infant mortality in New York City. A study of the results accomplished by infant-life saving agencies. New York, NY: The Rockefeller Foundation International Health Board, **1921**.
24. Morley DC. Measles in the developing world. Proc R Soc Med **1974**; 67:1112–5.
25. Koster FT, Curlin GC, Aziz KM, Haque A. Synergistic impact of measles and diarrhoea on nutrition and mortality in Bangladesh. Bull World Health Organ **1981**; 59:901–8.
26. Feachem RG, Koblinsky MA. Interventions for the control of diarrhoeal diseases among young children: measles immunization. Bull World Health Organ **1983**; 61:641–52.
27. Zarocostas J. Mortality from measles fell by 91% in Africa from 2000 to 2006. BMJ **2007**; 335:1173.
28. Ramos-Alvarez M, Sabin AB. Enteropathogenic viruses and bacteria; role in summer diarrheal diseases of infancy and early childhood. J Am Med Assoc **1958**; 167:147–56.
29. Olarte J, Ramos-Alvarez M, Galindo E. Isolation of *Shigella*, *Salmonella* & coli enteropathogens in rectal swabs of 802 cases of sporadic diarrhea. Bol Med Hosp Infant Mex **1957**; 14:257–62.
30. Moore Ha, Delacruz E, Perez Fi. Prevalence of enteropathogenic bacteria in Costa Rica. Bol Oficina Sanit Panam **1964**; 56:564–73.
31. Malherbe H, Roux P, Kahn E. The role of enteropathogenic bacteria and viruses in acute diarrhoeal disorders of infancy and childhood in Johannesburg. ii. "Non-specific" gastro-enteritis. S Afr Med J **1963**; 37:259–61.
32. Roux P, Kahn E, Malherbe H, Cassel R. The role of enteropathogenic bacteria and viruses in acute diarrhoeal disorders of infancy and childhood in Johannesburg. i. Summer diarrhoea. S Afr Med J **1963**; 37:256–9.
33. Hodes HL. The etiology of infantile diarrhea. Adv Pediatr **1956**; 8:13–52.
34. Dean AG, Ching YC, Williams RG, Harden LB. Test for *Escherichia coli* enterotoxin using infant mice: application in a study of diarrhea in children in Honolulu. J Infect Dis **1972**; 125:407–11.
35. Ryder RW, Sack DA, Kapikian AZ, et al. Enterotoxigenic *Escherichia coli* and reovirus-like agent in rural Bangladesh. Lancet **1976**; 1:659–63.
36. Tee GH, Moody AH, Cooke AH, Chiodini PL. Comparison of techniques for detecting antigens of *Giardia lamblia* and *Cryptosporidium parvum* in faeces. J Clin Pathol **1993**; 46:555–8.
37. Moodley D, Jackson TF, Gathiram V, van den Ende J. A comparative assessment of commonly employed staining procedures for the diagnosis of cryptosporidiosis. S Afr Med J **1991**; 79:314–7.
38. Shahid NS, Rahman AS, Anderson BC, Mata LJ, Sanyal SC. Cryptosporidiosis in Bangladesh. Br Med J (Clin Res Ed) **1985**; 290:114–5.
39. Thompson MR, Jordan RL, Luttrell MA, et al. Blinded, two-laboratory comparative analysis of *Escherichia coli* heat-stable enterotoxin production by using monoclonal antibody enzyme-linked immunosorbent assay, radioimmunoassay, suckling mouse assay, and gene probes. J Clin Microbiol **1986**; 24:753–8.
40. Lanata CF, Kaper JB, Baldini MM, Black RE, Levine MM. Sensitivity and specificity of DNA probes with the stool blot technique for detection of *Escherichia coli* enterotoxins. J Infect Dis **1985**; 152:1087–90.
41. Levine MM, Prado V, Robins-Browne R, et al. Use of DNA probes and HEp-2 cell adherence assay to detect diarrheagenic *Escherichia coli*. J Infect Dis **1988**; 158:224–8.
42. Gicquelais KG, Baldini MM, Martinez J, et al. Practical and economical method for using biotinylated DNA probes with bacterial colony blots to identify diarrhea-causing *Escherichia coli*. J Clin Microbiol **1990**; 28:2485–90.
43. Levine MM, Ferreccio C, Prado V, et al. Epidemiologic studies of *Escherichia coli* infections in a low socioeconomic level periurban community in Santiago, Chile. Am J Epidemiol **1993**; 138:849–69.
44. Vidal M, Kruger E, Duran C, et al. Single multiplex PCR assay to identify simultaneously the six categories of diarrheagenic *Escherichia coli* associated with enteric infections. J Clin Microbiol **2005**; 43:5362–5.
45. Vidal R, Vidal M, Lagos R, Levine M, Prado V. Multiplex PCR for diagnosis of enteric infections associated with diarrheagenic *Escherichia coli*. J Clin Microbiol **2004**; 42:1787–9.
46. Rappelli P, Folgosa E, Solinas ML, et al. Pathogenic enteric *Escherichia coli* in children with and without diarrhea in Maputo, Mozambique. FEMS Immunol Med Microbiol **2005**; 43:67–72.
47. Newman RD, Jaeger KL, Wuhib T, Lima AA, Guerrant RL, Sears CL. Evaluation of an antigen capture enzyme-linked immunosorbent assay for detection of *Cryptosporidium* oocysts. J Clin Microbiol **1993**; 31:2080–4.
48. Youn S, Kabir M, Haque R, Petri WA Jr. Evaluation of a screening test for detection of *Giardia* and *Cryptosporidium* parasites. J Clin Microbiol **2009**; 47:451–2.
49. Mubashir M, Khan A, Baqai R, et al. Causative agents of acute diarrhoea in the first 3 years of life: hospital-based study. J Gastroenterol Hepatol **1990**; 5:264–70.
50. Chatterjee BD, Thawani G, Sanyal SN. Etiology of acute childhood diarrhoea in Calcutta. Trop Gastroenterol **1989**; 10:158–66.
51. Casalino M, Yusuf MW, Nicoletti M, et al. A two-year study of enteric infections associated with diarrhoeal diseases in children in urban Somalia. Trans R Soc Trop Med Hyg **1988**; 82:637–41.
52. Simango C, Dindiwe J. The aetiology of diarrhoea in a farming community in Zimbabwe. Trans R Soc Trop Med Hyg **1987**; 81:552–3.
53. Nathoo KJ, Mason PR, Trijssenaar FJ, Lyons NF, Tswana SA. Microbial pathogens associated with diarrhoea in children admitted to Harare Hospital for rehydration therapy. Cent Afr J Med **1986**; 32:118–23.
54. Mackenjee MK, Coovadia YM, Coovadia HM, Hewitt J, Robins-Browne RM. Aetiology of diarrhoea in adequately nourished young African children in Durban, South Africa. Ann Trop Paediatr **1984**; 4:183–7.
55. Huilan S, Zhen LG, Mathan MM, et al. Etiology of acute diarrhoea among children in developing countries: a multicentre study in five countries. Bull World Health Organ **1991**; 69:549–55.
56. Khan MM, Iqbal J, Ghafoor A, Burney MI. Aetiologic agents of diarrhoeal diseases in hospitalised children in Rawalpindi, Pakistan. J Diarrhoeal Dis Res **1988**; 6:228–31.
57. Gascon J, Vargas M, Schellenberg D, et al. Diarrhea in children under 5 years of age from Ifakara, Tanzania: a case-control study. J Clin Microbiol **2000**; 38:4459–62.
58. Ogunsanya TI, Rotimi VO, Adenuga A. A study of the aetiological agents of childhood diarrhoea in Lagos, Nigeria. J Med Microbiol **1994**; 40:10–4.

59. Loening WE, Coovadia YM, van den Ende J. Aetiological factors of infantile diarrhoea: a community-based study. Ann Trop Paediatr **1989**; 9:248–55.
60. Haque R, Mondal D, Karim A, et al. Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. Clin Infect Dis **2009**; 48:1191–7.
61. Hasan KZ, Pathela P, Alam K, et al. Aetiology of diarrhoea in a birth cohort of children aged 0–2 year(s) in rural Mirzapur, Bangladesh. J Health Popul Nutr **2006**; 24:25–35.
62. Baqui AH, Sack RB, Black RE, et al. Enteropathogens associated with acute and persistent diarrhea in Bangladeshi children <5 years of age. J Infect Dis **1992**; 166:792–6.
63. Albert MJ, Faruque AS, Faruque SM, Sack RB, Mahalanabis D. Case-control study of enteropathogens associated with childhood diarrhea in Dhaka, Bangladesh. J Clin Microbiol **1999**; 37:3458–64.
64. Nakano T, Binka FN, Afari EA, et al. Survey of enteropathogenic agents in children with and without diarrhoea in Ghana. J Trop Med Hyg **1990**; 93:408–12.
65. Addy PA, Antepim G, Frimpong EH. Prevalence of pathogenic *Escherichia coli* and parasites in infants with diarrhoea in Kumasi, Ghana. East Afr Med J **2004**; 81:353–7.
66. Household KC, Dove MG, Smith MS. Enteropathogens associated with acute infantile diarrhoea in Bloemfontein, South Africa. J Trop Pediatr **1987**; 33:287–8.
67. Opintan JA, Newman MJ, Ayeh-Kumi PF, et al. Pediatric diarrhea in southern Ghana: etiology and association with intestinal inflammation and malnutrition. Am J Trop Med Hyg **2010**; 83:936–43.
68. Albert MJ, Faruque SM, Faruque AS, et al. Controlled study of *Escherichia coli* diarrheal infections in Bangladeshi children. J Clin Microbiol **1995**; 33:973–7.
69. Hoge CW, Echeverria P, Rajah R, et al. Prevalence of *Cyclospora* species and other enteric pathogens among children less than 5 years of age in Nepal. J Clin Microbiol **1995**; 33:3058–60.
70. Howard P, Alexander ND, Atkinson A, et al. Bacterial, viral and parasitic aetiology of paediatric diarrhoea in the highlands of Papua New Guinea. J Trop Pediatr **2000**; 46:10–4.
71. Stanton B, Silimperi DR, Khatun K, et al. Parasitic, bacterial and viral pathogens isolated from diarrhoeal and routine stool specimens of urban Bangladeshi children. J Trop Med Hyg **1989**; 92:46–55.
72. Anvikar AR, Dolla C, Dutta S, et al. Role of *Escherichia coli* in acute diarrhoea in tribal preschool children of central India. Paediatr Perinat Epidemiol **2008**; 22:40–6.
73. Molbak K, Wested N, Hojlyng N, et al. The etiology of early childhood diarrhea: a community study from Guinea-Bissau. J Infect Dis **1994**; 169:581–7.
74. Qadri F, Saha A, Ahmed T, Al TA, Begum YA, Svennerholm AM. Disease burden due to enterotoxigenic *Escherichia coli* in the first 2 years of life in an urban community in Bangladesh. Infect Immun **2007**; 75:3961–8.
75. Valentiner-Branth P, Steinsland H, Fischer TK, et al. Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. J Clin Microbiol **2003**; 41:4238–45.
76. Georges MC, Wachsmuth IK, Meunier DM, et al. Parasitic, bacterial, and viral enteric pathogens associated with diarrhea in the Central African Republic. J Clin Microbiol **1984**; 19:571–5.
77. De MP, Brasseur D, Hemelhof W, Kalala T, Butzler JP, Vis HL. Enteropathogenic agents in children with diarrhoea in rural Zaire. Lancet **1983**; 1:516–8.
78. Mikhail IA, Fox E, Haberberger RL Jr., Ahmed MH, Abbatte EA. Epidemiology of bacterial pathogens associated with infectious diarrhea in Djibouti. J Clin Microbiol **1990**; 28:956–61.
79. United Nations Children's Fund. State of the world's children 1990. New York, NY: Oxford University Press, **1990**.
80. Nimri LF, Elnasser Z, Batchoun R. Polymicrobial infections in children with diarrhoea in a rural area of Jordan. FEMS Immunol Med Microbiol **2004**; 42:255–9.
81. Schorling JB, Wanke CA, Schorling SK, McAuliffe JF, De Souza MA, Guerrant RL. A prospective study of persistent diarrhea among children in an urban Brazilian slum. Patterns of occurrence and etiologic agents. Am J Epidemiol **1990**; 132:144–56.
82. Na'was TE, Abo-Shehada MN. A study of the bacterial and parasitic causes of acute diarrhoea in northern Jordan. J Diarrhoeal Dis Res **1991**; 9:305–9.
83. Rosa AC, Mariano AT, Pereira AM, Tibana A, Gomes TA, Andrade JR. Enteropathogenicity markers in *Escherichia coli* isolated from infants with acute diarrhoea and healthy controls in Rio de Janeiro, Brazil. J Med Microbiol **1998**; 47:781–90.
84. Ming ZF, Xi ZD, Dong CS, et al. Diarrhoeal disease in children less than one year of age at a children's hospital in Guangzhou, People's Republic of China. Trans R Soc Trop Med Hyg **1991**; 85:667–9.
85. Fang GD, Lima AA, Martins CV, Nataro JP, Guerrant RL. Etiology and epidemiology of persistent diarrhea in northeastern Brazil: a hospital-based, prospective, case-control study. J Pediatr Gastroenterol Nutr **1995**; 21:137–44.
86. Zaki AM, DuPont HL, el Alamy MA, et al. The detection of enteropathogens in acute diarrhea in a family cohort population in rural Egypt. Am J Trop Med Hyg **1986**; 35:1013–22.
87. Poocharoen L, Bruin CW, Sirisanthana V, Vannareumol P, Leechanachai P, Sukhavat K. The relative importance of various enteropathogens as a cause of diarrhoea in hospitalized children in Chiang Mai, Thailand. J Diarrhoeal Dis Res **1986**; 4:10–5.
88. Adkins HJ, Escamilla J, Santiago LT, Ranoa C, Echeverria P, Cross JH. Two-year survey of etiologic agents of diarrheal disease at San Lazaro Hospital, Manila, Republic of the Philippines. J Clin Microbiol **1987**; 25:1143–7.
89. Bodhidatta L, McDaniel P, Sornsakrin S, Srijan A, Serichantalergs O, Mason CJ. Case-control study of diarrheal disease etiology in a remote rural area in Western Thailand. Am J Trop Med Hyg **2010**; 83:1106–9.
90. Echeverria P, Taylor DN, Lexsomboon U, et al. Case-control study of endemic diarrheal disease in Thai children. J Infect Dis **1989**; 159:543–8.
91. Hien BT, Trang do T, Scheutz F, Cam PD, Molbak K, Dalsgaard A. Diarrhoeagenic *Escherichia coli* and other causes of childhood diarrhoea: a case-control study in children living in a wastewater-use area in Hanoi, Vietnam. J Med Microbiol **2007**; 56(Pt 8): 1086–96.
92. al-Bwardy MA, Ramia S, al-Frayh AR, et al. Bacterial, parasitic and viral enteropathogens associated with diarrhoea in Saudi children. Ann Trop Paediatr **1988**; 8:26–30.
93. Soltan Dallal MM, Moezardalan K. *Aeromonas* spp associated with children's diarrhoea in Tehran: a case-control study. Ann Trop Paediatr **2004**; 24:45–51.
94. Orlandi PP, Magalhaes GF, Matos NB, et al. Etiology of diarrheal infections in children of Porto Velho (Rondonia, Western Amazon region, Brazil). Braz J Med Biol Res **2006**; 39:507–17.
95. Paniagua GL, Monroy E, Garcia-Gonzalez O, Alonso J, Negrete E, Vaca S. Two or more enteropathogens are associated with diarrhoea in Mexican children. Ann Clin Microbiol Antimicrob **2007**; 6:17.
96. Al-Gallas N, Bahri O, Bouratbeen A, Ben HA, Ben AR. Etiology of acute diarrhea in children and adults in Tunis, Tunisia, with emphasis on diarrheagenic *Escherichia coli*: prevalence, phenotyping, and molecular epidemiology. Am J Trop Med Hyg **2007**; 77:571–82.
97. Lima AA, Moore SR, Barboza MS Jr., et al. Persistent diarrhea signals a critical period of increased diarrhea burdens and nutritional shortfalls: a prospective cohort study among children in northeastern Brazil. J Infect Dis **2000**; 181:1643–51.

98. Oyofo BA, Subekti D, Tjaniadi P, et al. Enteropathogens associated with acute diarrhea in community and hospital patients in Jakarta, Indonesia. FEMS Immunol Med Microbiol **2002**; 34:139–46.
99. Black RE, Lopez de Romana G, Brown KH, Bravo N, Bazalar OG, Kanashiro HC. Incidence and etiology of infantile diarrhea and major routes of transmission in Huascar, Peru. Am J Epidemiol **1989**; 129:785–99.
100. Cravioto A, Reyes RE, Ortega R, Fernandez G, Hernandez R, Lopez D. Prospective study of diarrhoeal disease in a cohort of rural Mexican children: incidence and isolated pathogens during the first two years of life. Epidemiol Infect **1988**; 101:123–34.
101. Meng CY, Smith BL, Bodhidatta L, et al. Etiology of diarrhea in young children and patterns of antibiotic resistance in Cambodia. Pediatr Infect Dis J **2011**; 30:331–5.

SUPPLEMENT ARTICLE

Some Epidemiologic, Clinical, Microbiologic, and Organizational Assumptions That Influenced the Design and Performance of the Global Enteric Multicenter Study (GEMS)

Tamer H. Farag, Dilruba Nasrin, Yukun Wu, Khitam Muhsen, William C. Blackwelder, Halvor Sommerfelt, Sandra Panchalingam, James P. Nataro, Karen L. Kotloff, and Myron M. Levine

Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

The overall aim of the Global Enteric Multicenter Study–1 (GEMS-1) is to identify the etiologic agents associated with moderate-to-severe diarrhea (MSD) among children <5 years of age, and thereby the attributable pathogen-specific population-based incidence of MSD, to guide investments in research and public health interventions against diarrheal disease. To accomplish this, 9 core assumptions were vetted through widespread consultation: (1) a limited number of etiologic agents may be responsible for most MSD; (2) a definition of MSD can be crafted that encompasses cases that might otherwise be fatal in the community without treatment; (3) MSD seen at sentinel centers is a proxy for fatal diarrheal disease in the community; (4) matched case/control is the appropriate epidemiologic design; (5) methods across the sites can be standardized and rigorous quality control maintained; (6) a single 60-day postenrollment visit to case and control households creates mini-cohorts, allowing comparisons; (7) broad support for GEMS-1 messages can be achieved by incorporating advice from public health spokespersons; (8) results will facilitate the setting of investment and intervention priorities; and (9) wide acceptance and dissemination of the GEMS-1 results can be achieved.

The Global Enteric Multicenter Study–1 (GEMS-1) aims to identify the etiologic agents associated with moderate-to-severe diarrhea (MSD) in 3 pediatric age groups (0–11, 12–23, and 24–59 months of age) and estimate the population-based incidence in 4 sites in sub-Saharan Africa and 3 sites in South Asia. It also intends to identify water/sanitation/hygiene risk factors that may favor the transmission of specific enteric pathogens or factors that may be protective against pathogens that cause MSD. Each GEMS-1 site has an associated defined population under prospective demographic surveillance.

Correspondence: Tamer H. Farag, PhD, Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201 (tfarag@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S225–31

DOI: 10.1093/cid/cis787

In the first article in this supplement, Levine et al provide the rationale for the GEMS-1 study and its genesis. In this article, we describe certain assumptions that were either fundamental in the design of GEMS-1, critical to its successful implementation, or necessary for the interpretation, widespread dissemination, and advocacy of its results.

ASSUMPTION 1: A LIMITED NUMBER OF ETIOLOGIC AGENTS MAY BE RESPONSIBLE FOR A DISPROPORTIONATELY LARGE FRACTION OF ALL MSD

The genesis of GEMS-1 was largely driven by the fundamental concept that it may be possible to diminish the morbidity and mortality burden caused by diarrheal disease in young children in developing countries if effective vaccines were available and introduced

against the main etiologic agents that cause potentially fatal MSD. As reviewed by Levine et al (this supplement), many infectious agents have been purported to have the ability to cause severe and fatal diarrheal disease in young children. If the overall burden of diarrheal disease represents the collective consequence of a large number of pathogens, each making a small contribution, it is unlikely that the vaccine approach would be feasible from the epidemiologic, industrial, financial, or public health perspectives. On the other hand, if only a relatively small number of etiologic agents in infants 0–11 months, toddlers 12–23 months, and preschool children (24–59 months of age) were found to be responsible for a notable proportion of all MSD, then a vaccine-based strategy could be feasible. In undertaking GEMS-1, we were following the assumption that if we could identify a limited number of target pathogens for antidiarrheal vaccine development and/or delivery, a small number of vaccines could protect against a large fraction of the multicausal syndrome known as diarrheal disease. And if those vaccines were implemented widely in the developing world, the slope of decrease of the global morbidity and mortality burden from MSD could be steepened [1].

ASSUMPTION 2: A DEFINITION OF ACUTE MSD CAN BE CRAFTED THAT LIKELY ENCOMPASSES THOSE CASES THAT MIGHT BE FATAL IF THEY WERE NOT READILY TREATED

Fundamental to the GEMS-1 project was creating a definition for MSD that would receive wide acceptance by clinicians and epidemiologists and that would be practical for use in the field and applicable in both sub-Saharan Africa and South Asian settings. As described by Kotloff et al (this supplement), this took considerable discussion among GEMS investigators and clinical consultants on the Steering Committee on Epidemiologic and Clinical Issues (Table 1) who shared their experiences and constraints. Through consensus, a useful definition of MSD was successfully crafted: a child with diarrhea (≥3 abnormally loose stools) within the previous 24 hours with onset within the previous 7 days, following at least 7 days without diarrhea, and accompanied by evidence of clinically significant dehydration (loss of skin turgor, sunken eyes, or a decision by the clinician to administer intravenous fluids), dysentery (blood in the stool), or a clinical decision to hospitalize the

Table 1. Steering Committee on Epidemiologic and Clinical Issues

Members	Affiliation
External Members	
Fred N. Binka	University of Ghana
John D. Clemens	International Vaccine Institute; University of California, Los Angeles (currently)
Dani I. Cohen	Tel Aviv University
Roger I. Glass	Fogarty International Center
Halvor Sommerfelt	University of Bergen
Paul D. Stolley	University of Maryland School of Medicine
Internal Members	
Richard A. Adegbola	Medical Research Council, Gambia
Adebayo Akinsola	Medical Research Council, Gambia
Pedro L. Alonso	University of Barcelona
Sujit K. Battacharya	National Institute of Cholera and Enteric Diseases; Indian Council of Medical Research; World Health Organization (currently)
Robert F. Breiman	Centers for Disease Control and Prevention
Sumon K. Das	International Centre for Diarrhoeal Disease Research, Bangladesh
Abu S. Faruque	International Centre for Diarrhoeal Disease Research, Bangladesh
Philip C Hill	Medical Research Council, Gambia
Byomkesh Manna	National Institute of Cholera and Enteric Diseases
Eric D. Mintz	Centers for Disease Control and Prevention
Tacilta Nhampossa	Centro de Investigaçao em Saude da Manhiça
Richard Omore	Centers for Disease Control and Prevention
Ciara O'Reilly	Centers for Disease Control and Prevention
Debasish Saha	Medical Research Council, Gambia
Samba O. Sow	Center for Vaccine Development, Mali
Dipika Sur	National Institute of Cholera and Enteric Diseases
Anita K. M. Zaidi	Aga Khan University

child. Thus, the case definition encompasses episodes that might be fatal based on dehydration, dysentery, or systemic toxicosis. Cases progressing to persistent diarrhea, another clinical diarrheal syndrome associated with increased risk of fatality, were detected by means of a simple 14-day follow-up visual aid by which the mother or other caretaker recorded the number of days that diarrhea continued unabated (see Kotloff et al, this supplement). Following completion of the initial 3-year case control study of the etiology and burden of MSD (thereafter referred to as "GEMS-1"), a one-year, carry-on study called "GEMS-1A" was initiated to study diarrhea cases not meeting the definition for MSD. These less severe diarrhea (LSD) cases are being enrolled alongside MSD, with controls defined identically for both groups, enabling comparison of etiology and other outcomes between MSD and LSD.

ASSUMPTION 3: MSD SEEN AT SENTINEL HOSPITALS AND HEALTH CENTERS IS A PROXY FOR FATAL PEDIATRIC DIARRHEAL DISEASE IN THE COMMUNITY

It was necessary to have a strategy to capture cases of MSD in a practical way. The assumption made by the designers of the GEMS-1 project is that cases of diarrheal disease for which healthcare is sought at hospitals or health centers are likely to be more severe than cases that might be detected at the household level. Thus, the assumption was that passive surveillance at sentinel fixed healthcare facilities would be enriched for MSD cases. Accordingly, the selection of the most appropriate sentinel healthcare facilities was a critical step in the GEMS-1 design. As described by Kotloff et al, a large baseline random survey (Health Care Utilization and Attitudes Survey [HUAS]) carried out within the linked defined population generated the necessary information to identify the most appropriate facilities. This survey also provided information on healthcare preferences to allow adjustment so that population-wide incidence rates could be generated based on extrapolation of data gathered from sentinel health centers (SHCs).

ASSUMPTION 4: THE APPROPRIATE EPIDEMIOLOGIC DESIGN FOR IDENTIFYING THE RELATIVE IMPORTANCE OF PATHOGENS ASSOCIATED WITH MSD IS A MATCHED CASE/CONTROL STUDY

Whereas diarrheal illness among children in developing countries is common, cases of a severity that meet the definition of MSD constitute only a fraction of all pediatric diarrhea cases. Thus, utilizing a prospective cohort design with active household-based surveillance would require a very large cohort. Moreover, active household surveillance would likely modify the true incidence of MSD, since milder forms of diarrhea (when detected) would be treated and the evolution of the illness might be interrupted; that is, without having been

Table 2. Steering Committee on Microbiological Issues

Members	Affiliation
External Members	
Roger I. Glass	Fogarty International Center
Patrick R. Murray	National Institutes of Health; BD Diagnostics (currently)
Philippe J. Sansonetti	Institut Pasteur
Duncan A. Steele	World Health Organisation; PATH; Bill and Melinda Gates Foundation (currently)
Internal Members	
Martin Antonio	Medical Research Council
Anowar Hossain	International Centre for Diarrhoeal Diseaese Research, Bangladesh
Eric R. Houpt	University of Virginia
Inacio M. Mandomando	Centro de Investigaçao em Saude da Manhiça
Benjamin Ochieng	Centers for Disease Control and Prevention
Joseph Oundo	Centers for Disease Control and Prevention
William A. Petri	University of Virginia
Valeria Prado	University of Chile
T. Ramamurthy	National Institute of Cholera and Enteric Diseases
Ann-Mari Svennerholm	University of Göteborg
Boubou Tamboura	Center for Vaccine Development, Mali
Roberto Vidal	University of Chile
Anita K. M. Zaidi	Aga Khan University

detected by household visits, some of those diarrhea cases would have progressed to MSD and would have been detected when care was sought at a fixed healthcare facility. By contrast, a properly designed and executed prospective case/control study can accomplish the objective in a more practical, economical, and cost-effective way. An age-, sex- and village or neighborhood-matched control selection strategy was selected to limit potential confounding by factors that may not be easily controlled for in statistical analysis, thus ensuring the integrity of the results. The advantages and caveats of case/control studies have been reviewed [2, 3]. This type of study must be meticulously designed both with respect to identification of the MSD cases and selection of matched controls. A Steering Committee on Epidemiologic and Clinical Issues replete with individuals with expertise in case/control study design and in the performance of such studies in developing countries (Table 1) was instrumental in influencing the design of GEMS-1 and in attention to details of both case and control selection.

ASSUMPTION 5: THE CLINICAL AND LABORATORY METHODS ACROSS THE SITES COULD BE STANDARDIZED AND GOOD CLINICAL PRACTICES, GOOD CLINICAL LABORATORY PRACTICES, AND RIGOROUS QUALITY CONTROL COULD BE MAINTAINED THROUGHOUT THE STUDY

There exist institutions in the developing world such as the International Center for Diarrhoeal Diseases Research, Bangladesh (ICDDR,B) in Dhaka and the National Institute for Cholera and Enteric Diseases (NICED) in Kolkata, India, that have exceptional laboratory infrastructure and expertise for the detection of diarrheal pathogens and their characterization using sophisticated techniques. However, in very high-mortality areas of sub-Saharan Africa, there are no comparable venerable institutions with similar track records of sophisticated laboratory competence in detecting and characterizing the broad range of viral, bacterial and protozoal pathogens. On the other hand, there were potential field sites in areas of high young child mortality in sub-Saharan Africa with more general laboratory infrastructure, including where sophisticated techniques were used for the detection of other pathogens. We made the assumption that the capacity of these laboratories could be strengthened and expanded so that they could utilize the wide array of diagnostic technologies (including multiplex polymerase chain reaction) described by Panchalingam et al in this supplement for the detection of a very wide range of diarrheal pathogens.

Of equal importance, the new techniques not only had to be introduced but had to be standardized across all the GEMS sites and a system for continuous quality control had to be assured. The intention was to accomplish this with intensive initial training workshops, additional on-site training and by periodic site visits by an external laboratory supervisor with expertise in quality control and experience working in laboratories in the developing world. At all sites, the GEMS clinical microbiology procedures follow the tenets of Good Clinical Laboratory Practices.

ASSUMPTION 6: MAKING A SINGLE 60-DAY POSTENROLLMENT VISIT TO CASE AND CONTROL HOUSEHOLDS CREATES PROSPECTIVE MINI-COHORTS FOR COMPARISONS

Some diarrheal pathogens can lead to death by dehydration stemming from losses of body water and electrolytes through loose stools and vomiting, and accompanied by diminished intake [4, 5]. When these abnormal losses and diminished intake occur in an infant (whose daily water and electrolyte requirements per kilogram greatly exceed those of an older child or adult), and if prompt and adequate rehydration therapy is not introduced, severe dehydration, renal shutdown, and death can ensue [4, 5]. Some enterotoxigenic bacterial pathogens such as *Vibrio cholerae* O1 and some strains of enterotoxigenic *Escherichia coli* can cause such severe purging that even an older child can become severely dehydrated. However, once prompt and appropriate therapy is administered, survival of the patient is virtually assured and within a few days they are back to normal. In contrast, some enteric pathogens that cause pediatric diarrhea invade the intestinal mucosa and cause much destruction accompanied by

Table 3. Steering Committee on Nutritional Issues

Member	Affiliation
Claudio E. Lanata	Instituto de Investigacíon Nutricional
Reynaldo Martorell	Emory University
Rebecca J. Stoltzfus	Cornell University

Table 4. Steering Committee on Biostatistical Issues

Member	Affiliation
Barry I. Graubard	National Cancer Institute
Lawrence H. Moulton	Johns Hopkins University
William K. Pan	Johns Hopkins University
Peter Smith	London School of Tropical Medicine and Hygiene
Janet Wittes	Statistics Collaborative

Table 5. International Strategic Advisory Committee

Member	Affiliation
George E. Armah	University of Ghana
Zulfiqar A. Bhutta[a]	Aga Khan University
Fred N. Binka[a]	University of Ghana
Robert E. Black	Johns Hopkins University
A. Louis Bourgeois	PATH
Philip J. Cooper	Liverpool School of Tropical Medicine
Alejandro Cravioto	International Centre for Diarrhoeal Disease Research, Bangladesh
Valerie A. Curtis	London School of Hygiene and Tropical Medicine
Gordon Dougan	Wellcome Trust Sanger Institute
Kenneth C. Earhart	Centers for Disease Control and Prevention
Adenike Grange	Independent
George E. Griffin[b]	St George's, University of London
Gangadeep Kang	Christian Medical College
Claudio E. Lanata	Instituto de Investigacíon Nutricional
Reynaldo Martorell	Emory University
G. Balakrish Nair	National Institute of Cholera and Enteric Diseases
Miguel O'Ryan	University of Chile
Philippe J. Sansonetti	Pasteur Institute
Peter Smith	London School of Tropical Medicine and Hygiene

[a] Co-chair.

[b] Principal co-chair.

neutrophilic infiltration. *Shigella* is the prototype for this type of pathogen, as the mucosal destruction can result in outright bloody diarrhea (ie, dysentery) [6–8]. Other enteric pathogens such as enteropathogenic *E. coli* (EPEC) and *Cryptosporidium* cause other forms of mucosal disruption and modification characterized by effacement of enterocytes and pedestal formation [9, 10] or an unusual form of shallow invasion [11]. With pathogens such as *Shigella*, EPEC, and *Cryptosporidium* that result in striking pathologic changes in the intestinal mucosa, one cannot help but wonder whether over a more extended period of observation some episodes of diarrheal illness may result in adverse nutritional consequences and death many days beyond the initial acute injury, but by which time the subject may no longer have overt diarrhea. Presumably, cases of diarrheal illness accompanied by significant mucosal destruction or modification may be at particular risk of delayed consequences.

An important component of GEMS is a large case/control study. Onto that platform we have inserted the performance of a single visit to the households of cases and controls approximately 60 days after enrollment. At that time, follow-up anthropometric measurements are made (in particular, length/height) to determine whether the linear growth of cases over that period are the same or different from their matched controls, and whether growth impediments were associated with infection with specific enteropathogens. At the time of

Figure 1. Organogram for the Global Enteric Multicenter Study-1 (GEMS-1). This flowchart shows the organizational structure between the GEMS-1 coordinating center leadership, based at the Center for Vaccine Development, and the site principal investigators, based on the ground in the 7 country sites. It also shows the roles played by the various advisory bodies. Abbreviations: CVD, Center for Vaccine Development; PI, principal investigator.

this 60-day follow-up visit, we are also ascertaining whether the child (whether case or control) is still alive. Our assumption is that the risk of death may be greater over that approximately 60-day period for the cases than for their matched controls. If such an effect is discovered, intensive analyses of the data would be undertaken to attempt to discover important determinants, including baseline differences, host risk factors, and possible associations with specific pathogens.

ASSUMPTION 7: BROAD SUPPORT AND "BUY-IN" TO THE CONCEPT OF GEMS-1 CAN BE ACHIEVED BY INCORPORATING THE BEST ADVICE FROM MANY MEMBERS OF THE PUBLIC HEALTH, MICROBIOLOGIC, EPIDEMIOLOGIC, PEDIATRIC, NUTRITION SCIENCE, BIOSTATISTICAL, AND DISEASE CONTROL COMMUNITIES IN VARIOUS PARTICIPATORY AND ADVISORY ROLES

The GEMS is a historical, complex, highly ambitious multicenter project that is unlikely to be repeated. Therefore, it is critical that there be wide conceptual support for the concept and methods. Since the results of GEMS will indicate that some pathogens appear to be more important than others for MSD, champions of one or another pathogen may be disappointed by the results even as other champions are elated by the recognition of how their pathogen of interest has fared. Therefore, prior to the initiation of the studies themselves, it was critical that the epidemiologic, clinical microbiologic, data management, and statistical analysis methods be vetted by a broad array of experts and stakeholders. For this reason, the process of designing the GEMS clinical protocol, and selection of the pathogens to be detected and with what laboratory methods, was painstaking and iterative, and was accomplished with the assistance of the Steering Committee on Epidemiologic and Clinical Issues (Table 1) and a Steering Committee on Microbiological Issues (Table 2). Advice and guidance on anthropometric and nutritional aspects of GEMS were provided by a Steering Committee on Nutritional Issues (Table 3), while the statistical strategies of analyzing the data were greatly influenced by a Steering Committee on Biostatistical Issues (Table 4). Each of these steering committees included widely recognized authorities who collectively provided broad expertise. Annual investigators' meetings were convened, either at GEMS sites in Africa (Mali and Mozambique) or Asia (India) or in the United States (Baltimore, Seattle, Philadelphia, and Boston). Some members of the various steering committees were invited to the investigators' meetings to provide liaison to their respective steering committees.

Once the collection of field and laboratory data were completed and data cleaning was under way toward achieving locked datasets, and once the analytical approaches had been agreed upon, a large independent external committee was formed, the GEMS International Strategic Advisory Committee (GEMS-ISAC; Table 5), to provide a fresh, very high-level overview of the project and to provide strategic advice. The idea for establishing the ISAC was conceived at the Bill & Melinda Gates Foundation by Thomas Brewer and Niranjan Bose. The ISAC, with Professor George Griffin of St George's Hospital Medical School as senior co-chair and with Dean Fred Binka of the University of Ghana and Professor Zulfqar Bhutta as the other two co-chairs, brings together approximately 20 internationally recognized leaders in public health, microbiology, epidemiology, nutrition, statistics, clinical infectious diseases, pediatrics, diarrheal diseases, environmental health (water/sanitation/hygiene), and environmental engineering, from both industrialized and developing countries.

The GEMS has been fortunate at all stages of the project, from conception to analyses of data, to have a series of program officers at the Bill & Melinda Gates Foundation who have been extraordinarily invaluable in supporting the project in every way possible. These include Thomas Brewer, Niranjan Bose, and Duncan Steele in the later years and Regina Rabinovich and Jan Agosti in the early years of the project.

An organogram of the management of the GEMS is shown in Figure 1. Also shown in this figure is the interface and point in the project the various committees interacted with the GEMS leadership and the GEMS investigators.

ASSUMPTION 8: RESULTS WILL FACILITATE THE SETTING OF INVESTMENT & INTERVENTION PRIORITIES

It is the hope and expectation of the GEMS investigators that the results of the GEMS will provide decision makers in international agencies, government development agencies, research funding agencies, philanthropic organizations, and public health implementers in developing countries an evidence base on the etiology and burden of more severe forms of diarrheal disease in developing countries, insights on MSD-associated mortality, and MSD-associated nutritional consequences.

The GEMS data on etiology of diarrheal disease in developing countries can also serve as a resource for groups that estimate mortality and disease burden (such as the World Health Organization [WHO] Child Health Epidemiology Reference Group, investigators at the Institute for Health Metrics and Evaluation, and similar groups), and groups that are concerned with food safety (such as the Foodborne Disease Burden Epidemiology Reference Group of WHO and the Food and Agriculture Organization) and groups preparing investment cases for diarrheal disease vaccines (eg, the *Shigella* and ETEC Investment Case efforts at PATH).

ASSUMPTION 9: WIDE ACCEPTANCE AND DISSEMINATION OF THE GEMS-1 RESULTS CAN BE ACHIEVED

Once the GEMS data have been fully analyzed and initial publications prepared, it is critical to disseminate short summaries of the most salient points and press releases in simple transparent language to convey the results to lay audiences, community leaders, politicians, trend setters, opinion makers, and the general population. Principal investigators at many of the GEMS sites have already begun to sensitize political leaders in their government of the importance of the GEMS data. For example, since rotavirus vaccines are expected to be introduced into the routine infant immunization schedule of the Expanded Programme on Immunization for a number of high-mortality developing countries in sub-Saharan Africa and Asia, the GEMS data can be exceedingly helpful to assist advocacy efforts.

Notes

Financial support. This study was funded by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Levine MM. Enteric infections and the vaccines to counter them: future directions. Vaccine **2006**; 24:3865–73.
2. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet **2002**; 359:431–4.
3. Rothman KJ, Greenland S. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, **2008**.
4. Booth IW, Levine MM, Harries JT. Oral rehydration therapy in acute diarrhoea in childhood. J Pediatr Gastroenterol Nutr **1984**; 3:491–9.
5. Levine MM, Pizarro D. Advances in therapy of diarrheal dehydration: oral rehydration. Adv Pediatr **1984**; 31:207–34.
6. Sansonetti PJ. War and peace at mucosal surfaces. Nat Rev Immunol **2004**; 4:953–64.
7. Cossart P, Sansonetti PJ. Bacterial invasion: the paradigms of enteroinvasive pathogens. Science **2004**; 304:242–8.
8. Levine MM, DuPont HL, Formal SB, et al. Pathogenesis of *Shigella dysenteriae* 1 (Shiga) dysentery. J Infect Dis **1973**; 127:261–70.
9. Moon HW, Whipp SC, Argenzio RA, Levine MM, Giannella RA. Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. Infect Immun **1983**; 41:1340–51.
10. Rothbaum RJ, Partin JC, Saalfield K, McAdams AJ. An ultrastructural study of enteropathogenic *Escherichia coli* infection in human infants. Ultrastruct Pathol **1983**; 4:291–304.
11. Elliott DA, Clark DP. *Cryptosporidium parvum* induces host cell actin accumulation at the host-parasite interface. Infect Immun **2000**; 68:2315–22.

SUPPLEMENT ARTICLE

The Global Enteric Multicenter Study (GEMS) of Diarrheal Disease in Infants and Young Children in Developing Countries: Epidemiologic and Clinical Methods of the Case/Control Study

Karen L. Kotloff,[1,2] William C. Blackwelder,[2] Dilruba Nasrin,[2] James P. Nataro,[1,2,a] Tamer H. Farag,[2] Annemieke van Eijk,[3] Richard A. Adegbola,[4,b] Pedro L. Alonso,[5,6] Robert F. Breiman,[7] Abu Syed Golam Faruque,[8] Debasish Saha,[4] Samba O. Sow,[9] Dipika Sur,[10] Anita K. M. Zaidi,[11] Kousick Biswas,[12] Sandra Panchalingam,[2] John D. Clemens,[13,c] Dani Cohen,[14] Roger I. Glass,[15] Eric D. Mintz,[16] Halvor Sommerfelt,[17,18] and Myron M. Levine[1,2]

[1]Department of Pediatrics, [2]Department of Medicine, Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; [3]Child and Reproductive Health Group, Liverpool School of Tropical Medicine, University of Liverpool, United Kingdom; [4]Medical Research Council (United Kingdom) Unit, Fajara, The Gambia; [5]Centre de Recerca en Salut Internacional de Barcelona, Hospital Clinic/Universitat de Barcelona, Spain; [6]Centro de Investigação em Saúde de Manhiça, Mozambique; [7]Global Disease Detection Division, Kenya Office of the US Centers for Disease Control and Prevention, Nairobi; [8]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; [9]Centre pour le Développement des Vaccins, Bamako, Mali; [10]National Institute of Cholera and Enteric Diseases, Kolkata, India; [11]Department of Paediatrics and Child Health, The Aga Khan University, Karachi, Pakistan; [12]Department of Veterans Affairs, Cooperative Studies Program Coordinating Center, Perry Point, Maryland; [13]International Vaccine Institute, Seoul, Korea; [14]Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel; [15]Fogarty International Center, National Institutes of Health, Bethesda, Maryland; [16]Division of Foodborne, Waterborne, and Environmental Diseases, US Centers for Disease Control and Prevention, Atlanta, Georgia; [17]Centre for International Health, University of Bergen, and [18]Division of Infectious Disease Control, Norwegian Institute of Public Health, Oslo, Norway

Background. Diarrhea is a leading cause of illness and death among children aged <5 years in developing countries. This paper describes the clinical and epidemiological methods used to conduct the Global Enteric Multicenter Study (GEMS), a 3-year, prospective, age-stratified, case/control study to estimate the population-based burden, microbiologic etiology, and adverse clinical consequences of acute moderate-to-severe diarrhea (MSD) among a censused population of children aged 0–59 months seeking care at health centers in sub-Saharan Africa and South Asia.

Methods. GEMS was conducted at 7 field sites, each serving a population whose demography and healthcare utilization practices for childhood diarrhea were documented. We aimed to enroll 220 MSD cases per year from selected health centers serving each site in each of 3 age strata (0–11, 12–23, and 24–59 months), along with 1–3 matched community controls. Cases and controls supplied clinical, epidemiologic, and anthropometric data at enrollment and again approximately 60 days later, and provided enrollment stool specimens for identification and characterization of potential diarrheal pathogens. Verbal autopsy was performed if a child died. Analytic strategies will calculate the fraction of MSD attributable to each pathogen and the incidence, financial costs, nutritional consequences, and case fatality overall and by pathogen.

Conclusions. When completed, GEMS will provide estimates of the incidence, etiology, and outcomes of MSD among infants and young children in sub-Saharan Africa and South Asia. This information can guide development and implementation of public health interventions to diminish morbidity and mortality from diarrheal diseases.

[a]Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia.

[b]GlaxoSmithKline Biologicals, Global Medical Affairs, Wavre, Belgium.

[c]Center for Global Infectious Diseases, University of California, Los Angeles.

Correspondence: Karen L. Kotloff, MD, University of Maryland School of Medicine, Center for Vaccine Development, 685 W Baltimore St, HSF 480, Baltimore, MD 21201 (kkotloff@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S232–45

DOI: 10.1093/cid/cis753

There has been substantial progress toward meeting Millennium Development Goals for child survival during the past 2 decades, such that under-5 mortality rates have decreased in every developing region of the world. Nonetheless, the rates have fallen more sharply in wealthier areas [1], resulting in a large and growing share of deaths in the poorer developing regions. As further declines are made possible by expanding interventions that target the principal causes of death and focus on the most vulnerable children, the availability of accurate, up-to-date assessments at country levels becomes even more important to guide strategic planning and resource allocation. This is especially true for sub-Saharan Africa and South Asia, where 50% and 32%, respectively, of the estimated annual 7.6 million under-5 deaths are now concentrated and where current, systematically collected information on the burden and major causes of child death is lacking [2].

Diarrheal diseases continue to be major causes of childhood mortality in developing countries. The proportion of deaths attributed to diarrhea among children 1–59 months of age is estimated to be 25% in Africa and 31% in South Asia [3]. These estimates were calculated by abstracting studies published between 1980 and 2009 that utilized verbal autopsies (postmortem interviews of family members) to assign cause of death in representative populations. Statistical models were applied to derive estimates of diarrhea-specific mortality and to extrapolate across countries and regions. Modeled estimates of disease burden are an invaluable metric for assessing progress toward achieving health objectives and for estimating the impact of various interventions; however, this approach faces limitations imposed by the quality, scope, age, and consistency of the underlying data. Analyses of the causes of childhood death based on verbal autopsies are subject to misclassification [4, 5], and if they include studies performed over several decades, the results may not reflect the current situation. Without concomitant morbidity assessments, one cannot determine to what extent secular trends in declining disease-specific mortality represent lower disease incidence or diminished case fatality (which can have different determinants and respond to different interventions). Rigorously conducted, prospective, population-based studies can be used to strengthen modeled disease burden estimates [6–8]. Moreover, such studies are essential for providing the detailed information needed to design new and improved interventions to prevent and treat the most life-threatening and disabling episodes, which, in the case of diarrhea, would include knowledge about the etiology, risk factors, nutritional sequelae, and case fatality.

We conducted the Global Enteric Multicenter Study (GEMS), a 3-year, prospective, age-stratified, matched case/control study of moderate-to-severe diarrhea (MSD) among children 0–59 months of age belonging to a censused population and seeking care at hospitals and health centers at 7 sites located in sub-Saharan Africa and South Asia. A common research protocol with standardized epidemiologic and microbiologic methods was used to facilitate intersite comparisons and allow aggregate estimates of etiology and incidence. This paper describes the study design, including site selection, a surveillance system to characterize the demography and healthcare utilization practices of the catchment population, methods for enrollment, data collection and follow-up of case and control children, and quality control activities. We discuss challenges encountered in the implementation of a large study involving heterogeneous populations located in resource-poor settings.

METHODS

Objectives of the Study

The primary objective of GEMS was to measure the population-based burden, microbiologic etiology, and adverse clinical consequences of MSD in developing countries, overall and by age, pathogen, site, and clinical syndrome (simple nonbloody diarrhea, dysentery, or profuse watery diarrhea). The adverse clinical consequences of interest included growth faltering according to World Health Organization (WHO) standards [9], persistent diarrhea lasting ≥14 days, and death. The secondary objectives were (1) to determine the antigenic and genotypic characteristics of the leading pathogens to guide vaccine development; (2) to elucidate the risk factors attributable to the host, the microorganism, and the environment that are associated with the occurrence and adverse clinical outcomes of MSD; (3) to estimate the public and private financial costs, both direct and indirect, incurred during an episode of MSD; and (4) to create a central repository of well-characterized clinical specimens and isolated etiologic agents that can be shared with other investigators for future research.

Site Selection Criteria

Seven field sites were selected among countries in sub-Saharan Africa (Kenya, Mali, Mozambique, and The Gambia), and South Asia (Bangladesh, India, and Pakistan) with moderate to high under-five childhood mortality (Table 1). To create a broad view of enteric disease epidemiology, we chose sites that together exemplified a spectrum of child health indicators, with variations in the prevalence of malaria and human immunodeficiency virus (HIV) infection, and a mixture of urban, rural, and periurban settings (Table 1). Sites were required to have access to a population that had been or could undergo a census accompanied by an address system to allow households to be revisited in the future, and to 1 or more healthcare facilities that provide care to children from that population with diarrhea. Infrastructure with the potential for computerized data management, secure freezer storage, at

Table 1. Selected Child Health Indicators Available in 2005 and Used to Guide Site Selection[a]

							National Statistics								
												% Using Improved or Adequate[g]			
Country	City	Partner	Setting	No. SHCs in CCS	Population <5 y [c]	GNI per Capita (US$)	U5MR[b] (Country Rank)	% HIV+ (15–49 y)[d]	Malaria Rate[e]	% <5 y Wasted[f]	% <5 y Stunted[f]	Water	Sanitation	% <5 y Receiving ORS[h]	% 1 y DPT3[i]
Mali	Bamako	Centre pour le Développement des Vaccins du Mali (CVD-Mali)	Urban	9	31 768	290	220 (7)	1.9	62.2	11	38	76	59	45	69
The Gambia	Basse	Medical Research Council (MRC)	Rural	5	29 076	310	123 (37)	1.2	ND	9	19	77	46	38	90
Mozambique	Manhiça	Centro de Investigação em Saúde de Manhiça (CISM)	Rural	5	15 380	210	158 (24)	12.2	269.7	4	41	24	14	33	72
Kenya	Nyanza Province	CDC/Kenya Medical Research Institute (KEMRI) Research Station/CDC	Rural	11	21 603	390	123 (37)	6.7	3.9	6	31	46	43	15	73
India	Kolkata, W. Bengal	National Institute of Cholera and Enteric Diseases (NICED)	Urban	2	13 416	530	87 (54)	ND	1.7	16	46	96	58	22	70
Bangladesh	Mirzapur	International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B)	Rural	1	25 560	400	69 (62)	ND	0.4	10	45	72	39	35	85
Pakistan	Karachi (Bin Qasim Town)	Aga Khan University	Peri urban	7	25 659	470	103 (47)	0.1	0.8	13	37	87	35	33	67

Abbreviations: CCS, case/control study; CDC, Centers for Disease Control and Prevention; DPT3, complete coverage with diphtheria-pertussis-tetanus vaccine; GNI, gross national income; HIV, human immunodeficiency virus; ND, no data; ORS, oral rehydration solution; SHC, sentinel health centers where children with moderate-to-severe diarrhea were enrolled in the CCS; U5MR, under-5 mortality rate.

[a] All data pertain to 2003, with the exception of access to improved water and adequate sanitation, which pertain to 2002 [42], and the population <5 years, as described below.

[b] Value is calculated per 1000 live births and ranked out of 192 countries for 2003 [42].

[c] The population <5 years of age represents the median value from sequential demographic surveillance system rounds conducted during the case/control study.

[d] Prevalence of HIV (percentage) among 15- to 49-year-olds, as of end of 2003 [42].

[e] Standardized reported malaria rate per 1000 population, 2003 for all countries but Kenya (2002) [43].

[f] Percentage of children <5 years of age with wasting or stunting graded as moderate or severe [42].

[g] Data shown pertain to urban areas when the study site is urban sites and rural areas when the site is rural. Data for rural areas were considered most appropriate to represent the study site in Pakistan [42].

[h] Percentage of children <5 years of age with diarrhea receiving oral rehydration and continued feeding 1994–2003 [42].

[i] Percentage of children who received DPT3 by 1 year of age [42].

least intermittent internet transmission, and the ability to ship specimens and strains abroad had to be available, with capabilities to perform coprocultures, antigen-detection tests, and nucleic-acid based assays.

Establishing a Sampling Frame for the Case/Control Study and Selecting Health Centers for Case Recruitment

The census at each site will enable population-based estimates of the outcomes of interest. Each census was continually updated using a demographic surveillance system (DSS) in which the households were visited every 4–6 months to record pregnancies, births, deaths, and migrations in and out of the area. Between DSS visits, we enlisted a community reporter from each neighborhood to meet weekly with local leaders (religious figures, political representatives, and elders) and midwives to detect births and deaths among children 0–59 months of age. The reporter visited near-term pregnant women as an additional means of capturing births. Keeping the DSS current was necessary to maintain an accurate sampling frame from which to select matched community controls for the case/control study, and for the timely performance of verbal autopsies, as described below.

In preparation for the case/control study, we performed a Health Care Utilization and Attitudes Survey (HUAS). An age-stratified sample of approximately 1000 children aged 0–59 months per site randomly selected from each updated DSS dataset was visited at home, and parents/primary caretakers were asked whether their child had experienced diarrhea during the previous 14 days. If so, the presence of findings suggestive of MSD was solicited (sunken eyes, wrinkled skin, hospitalization, receipt of intravenous hydration, or dysentery), and source(s) of healthcare were recorded. These data were used to adjust the size of the DSS population at each site as necessary to contribute the requisite number of cases of MSD to each age stratum, and to select 1 or more "sentinel" health centers (SHCs) serving the DSS population at each site (Table 1) as venues for the case/control study based on their potential to capture MSD cases from the DSS.

During the second and third years of the case/control study, an abbreviated HUAS questionnnaire (designated "HUAS-lite") was administered to caretakers of approximately 1000 randomly selected children aged 0–59 months (age-stratified) approximately every 4 to 6 months in association with the DSS interviews. HUAS-lite data were used to refine the selection of SHCs, to estimate the extent to which children with MSD who seek care at SHC are representative of children with MSD in the DSS population (by comparing features of those who do and do not seek care), and to calculate the proportion of children with MSD who sought care at the SHCs at each site (r value) as a means of extrapolating the overall and pathogen-specific MSD episodes enumerated at the SHCs to derive the incidence estimates for the entire DSS population (see Blackwelder, et al, this supplement).

Process Development and Training

Paper case report forms (CRFs) were created and translated into the 4 languages spoken by the interviewers (English, French, Portuguese, and dual Dholuo and English) according to preference of the local study teams. Interviews were always conducted in the native language of the respondent. Initial versions of the CRFs were field tested, then modified as needed at a 4-day study development meeting attended by each site's senior clinical investigators and study coordinators.

A pilot case/control study was conducted for approximately 3 months followed by the full, 36-month case/control study. Before the pilot and full-study initiations, we conducted a 5-day training program at each site, using interactive adult learning techniques with group participation, role playing, small group practice sessions, and evaluations of competency. We compiled an interviewers' and a supervisors' manual of procedures which served as the basis for training sessions. The curriculum covered the principles of human subjects research and elements of good clinical practices [10], how to conduct an interview (eg, issues of privacy, building rapport with the respondent, and asking questions in a nonjudgmental way), perform a focused physical examination, collect, process, and transport stool specimens, and document observations of water and sanitation facilities. The meaning of each question and response choice was discussed. Terms were defined, using pictures and graphics whenever applicable. Participants practiced a standardized format for handwriting English letters and numbers to reduce frequency of errors in data recording and entry. The supervisors' manual and training focused on supportive supervision techniques, training, handling underperforming staff, quality management, tracking study activities, and performing oversight, spot checks, and reinterviews to ensure the validity and reliability of the data.

We enlisted the assistance of an experienced anthropometrist to train the clinical and field staff at 1 site in Asia and 1 site in Africa in obtaining length/height, weight, and mid-upper arm circumference (MUAC) measurements. These 3-day training sessions also served the purpose of training the Epidemiology Team from the core site at the Center for Vaccine Development (CVD), University of Maryland, Baltimore (K. L. K., D. N., and T. H. F.), who then conducted similar training at the remaining 5 sites. On the third day of training, 10 children (5 aged 0–23 months and 5 aged 24–59 months) participated in a standardization session in which each trainee performed 2 independent measurements of the length/height and MUAC of each child. Intrarater reliability and validity were calculated using the anthropometrist (and later the CVD epidemiologist) as the "gold standard."

A difference of >0.5 cm was considered unacceptable when comparing a trainee's 2 measurements of the same child or when comparing the trainee's and the gold-standard measurements of the same child. Trainees with unacceptable performance were retrained until competency was achieved. A gold-standard measurer was identified at each site to supervise field measurements and to train all staff including newly hired staff every 4–6 months. Technical error of measurement and average bias will be calculated to assess inter- and intraobserver reproducibility as well as validity of measurements [11].

Scientific Oversight

We recognized that controversies persist regarding the most appropriate case definitions and detection methods for studying diarrhea, and that the approach chosen would impact the resultant estimates of disease burden [12, 13]. Consequently, as part of a consensus-building process, we assembled a Steering Committee on Epidemiologic and Clinical Issues comprising the lead investigators from each site, and a multinational group of 6 experts in diarrheal disease. The committee vetted the clinical protocol at the initial meeting. Thereafter, the committee was convened annually and on an ad hoc basis as issues arose. Once the study was initiated, external experts were assembled to form a Steering Committee on Nutritional Issues and a Steering Committee on Biostatistical Issues to review the analysis plan and to provide guidance as issues arose. In the final months of the study, an International Strategic Advisory Committee was formed to critically review the methodology and results and to advise the funding agency about the significance of the findings to inform its strategic planning for the future (Farag et al, this supplement).

Case Definition of MSD and Other Study Outcomes

The initial step in eligibility screening was the selection of children who fulfilled the WHO definition of diarrhea (≥3 abnormally loose stools in 24 hours [12]). In subsequent steps we identified cases of MSD, the primary outcome of interest, intending to capture diarrheal illnesses that would not be expected to resolve spontaneously without medical intervention or without sequelae, because these illnesses constitute priorities for development of vaccines and other new or improved preventive and therapeutic strategies. We reasoned that episodes that would qualify as MSD fell into 2 general categories: (1) those accompanied by dehydration to a degree that the child's survival would likely depend on access to life-saving rehydration fluids, and (2) those with evidence of inflammatory destruction of the intestinal mucosa, thereby at increased risk for disabling sequelae (such as persistent diarrhea [14] and stunting [15]) or death [16, 17].

To capture children who had potentially life-threatening diarrheal dehydration, we adapted the WHO definition of dehydration to our case definition of MSD [18, 19], choosing the most objective signs (sunken eyes more than usual and slow or very slow recoil after an abdominal wall "skin pinch"). In addition, we included the determination by a healthcare provider that the severity of dehydration warranted administration of intravenous fluids. Although not part of our case definition, other signs of dehydration proposed by WHO were documented, including restlessness or irritability and drinking eagerly or appearing thirsty (considered to be signs of "some" dehydration), and lethargy, loss of consciousness, inability to drink, or drinking poorly (as signs of "severe" dehydration). During analysis we will explore the impact on the study findings of including these other signs of dehydration in the definition of MSD. We considered adopting as inclusion criteria elements of systems used widely to define severe illness in rotavirus vaccine trials [20–22]. However, many of the components, such as total duration and maximum severity of diarrhea, vomiting, and fever, can only be determined in retrospect when the episode is resolving or resolved, at which point the decision to include a child in GEMS would already have been made. Instead, our approach has been to collect this information for exploration during analysis.

To capture children with evidence of diarrheal diseases caused by inflammation and mucosal injury in the case definition, we enrolled children with dysentery. Because there is no marker to predict which cases of dysentery are likely to experience clinically significant intestinal damage, we included all children with diarrhea who passed at least 1 stool containing visible blood according to either the caretaker or the clinician. Finally, we included children with diarrhea who appeared sufficiently ill to prompt the healthcare provider to recommend overnight admission to the hospital.

We restricted enrollment to children with acute MSD (≤7 days' duration) to maximize the opportunity to identify the inciting pathogen and to collect new episodes that can be used together with DSS and HUAS data to estimate annual incidence rates. We defined an episode of diarrhea as days with diarrhea beginning after at least 7 diarrhea-free days and ending when diarrhea is not present for 7 days [23, 24]. Although the WHO definition of a new episode of diarrhea requires only 3 diarrhea-free days [12], we chose a longer interval (as have other investigators [12, 25]) to increase our margin of certainty that the episode was new, recognizing that this approach could underestimate the incidence of MSD.

Case Ascertainment

Cases of MSD were identified in SHCs (hospital, urgent care facilities, and community clinics) to capture those illnesses that are most severe and that collectively constitute a significant cost in healthcare services, and thus would be targeted for prevention by vaccines and other interventions (Figure 1).

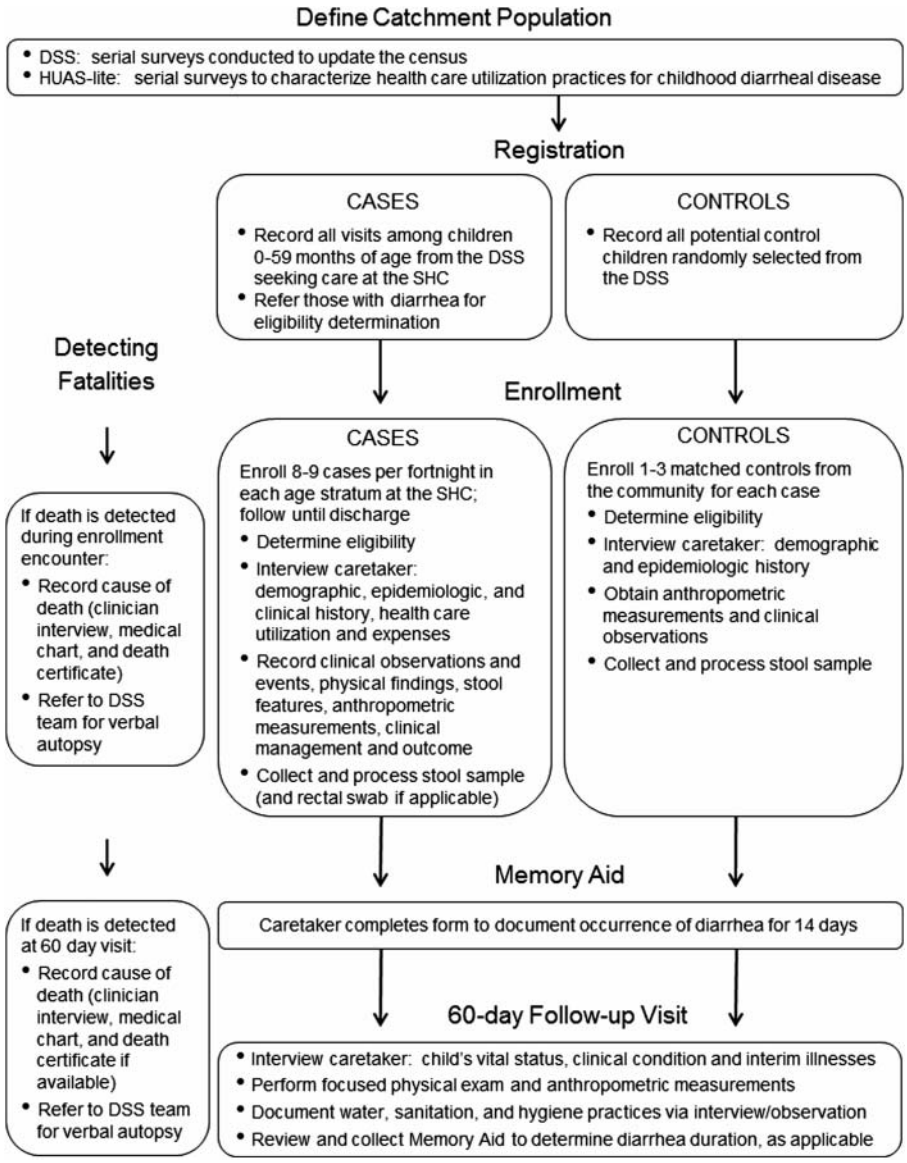

Figure 1. Flow diagram illustrating major study activities. Abbreviations: DSS, demographic surveillance system; SHC, sentinel health center; HUAS, Health Care Utilization and Attitudes Survey.

GEMS staff were situated in the intake area at each SHC to complete a registration log documenting each visit made by a child 0–59 months old belonging to the DSS. The GEMS registrar was given access to the DSS database to verify that a child belonged to the DSS, and to record each enrolled child's unique DSS number as a means for determining, at a later date, who was enrolled into GEMS more than once. Each visit was assigned a unique screening identification number, and the registrar recorded the date and time the child entered the SHC; the child's age, sex, and village/neighborhood; whether the child had diarrhea; and whether the child was hospitalized. The GEMS registrar referred all children from the DSS who were aged 0–59 months and had diarrhea to a GEMS clinician. The clinician informed the parent/primary caretaker about the study, determined the child's eligibility (Table 2), and obtained informed consent. If an eligible child was not enrolled, the reasons for nonenrollment were documented (eg, refusal, missed opportunity, stool sample inadequate or not obtained, 14-day quota filled, or child died before enrollment).

Each site aimed to enroll approximately 220 MSD patients per year into each of 3 age strata: 0–11 months, 12–23 months, and 24–59 months, totaling 1980 cases over 3 years. To ensure even sampling throughout the year, the target was to enroll approximately 8–9 cases per age stratum (25–26

Table 2. Inclusion Criteria for Cases

1. Child is 0–59 mo of age
2. Child belongs to the demographic surveillance system population at the site
3. Child is not currently enrolled as a case (meaning previously enrolled and pending 60-day visit)[a]
4. Child meets case definition of diarrhea (≥3 abnormally loose stools in the previous 24 h)
5. Diarrhea episode is: • Acute (onset within 7 d of study enrollment) and • Represents a new episode (onset after ≥7 diarrhea-free days) [23, 24]
6. Diarrhea is moderate-to-severe, meaning that the child met at least 1 of the following criteria: • Sunken eyes, confirmed by parent/primary caretaker as more than normal • Loss of skin turgor (determined by abdominal skin pinch (slow return [≤2 s] or very slow return [>2 s]) • Intravenous rehydration administered or prescribed • Dysentery (visible blood in a loose stool) • Hospitalized with diarrhea or dysentery

[a] A child was eligible to be enrolled as a case irrespective of whether he or she had been included as a case or as a control previously; whereas cases were eligible for reenrollment only after the 60-day follow-up visit had been completed, controls could be enrolled as a case at any time they met the criteria.

Table 3. Inclusion Criteria for Controls

1. Resides in demographic surveillance system area

2. Matched to the index case as follows:
• Age:
  ○ ±2 mo for cases 0–11 mo
  ○ ±4 mo for cases 12–59 mo
  ○ May not exceed the stratum boundaries of the case, eg, a control for an 11-mo-old case must be between the ages of 9 and 11 mo and a control for a 13-mo-old must be between the ages of 12 and 17 mo
• Same sex
• Residence: lives in the same or nearby village/neighborhood as the case[a]
• Time: enrolled within 14 d of presentation of the case

3. No diarrhea in the previous 7 d[b]

[a] Each site followed an algorithm beginning with the case's village/neighborhood, and then proceeding to villages/neighborhoods located at an increasing distance from the case's village/neighborhood until a control could be identified.

[b] Control children will be included in the analysis irrespective of whether they developed diarrhea after enrollment.

cases overall) per fortnight. This strategy prevented the strata from being filled prematurely in seasons with high volume and respected the capacity limitations of the clinical and microbiology personnel, but because all DSS children with MSD were recorded, temporal increases in the case load of MSD and of specific diarrheal pathogens could be measured. Analyses for events that might have seasonal variation will take into account the sampling fraction of MSD for each period.

Control Selection

For each child with MSD included in the study, we enrolled 1–3 control children without diarrhea from the DSS community (Figure 1) within 14 days of presentation of the index case. Sites tracked their ability to fill each age stratum on a fortnightly basis and followed an algorithm to determine the number of controls to enroll: 1:1 case:control matching if 7–9 cases were enrolled; 1:2 matching if 4–6 cases were enrolled, and 1:3 matching if ≤3 cases were enrolled. At least 4 children who met the matching criteria (Table 3) were randomly selected from the DSS database as potential controls. A field worker visited the home of selected children sequentially and explained all aspects of the study. If the parent/primary caretaker expressed interest and the child met eligibility criteria (Table 3), informed, written consent was obtained and arrangements were made to collect a stool sample, as described below. Reasons for not enrolling a selected child were documented (eg, refusal, not found at home after 3 attempts to contact, or failed to produce an adequate stool sample).

Data Collection at Enrollment From Cases and Controls

Case enrollment interviews took place at the SHC whereas control caretakers were interviewed at home. To facilitate linkage of our results with existing databases, we designed our caretaker interviews to include questions found in the primary sources of population-based data used to estimate child mortality in developing countries, such as the UNICEF-supported Multiple Indicator Cluster Surveys and the US Agency for International Development–supported Demographic and Heath Surveys [1]. Demographic information collected about the case or control and his/her household (defined as a group of people who share a cooking fire) included maternal education and household size (including the number of children <5 years old). Building materials and household possessions were documented (to assess potential risk factors for illness and as indicators for constructing a wealth index for each site [26]). Questions addressed handwashing practices and access and availability of improved water and sanitation facilities [27], animals on the premises, water treatment, sharing sanitation facilities, and disposal of the child's feces. The caretakers were queried about the child's clinical signs and symptoms; how the illness was managed prior to the SHC visit (the reference point was the current illness for cases and the most recent diarrheal illness for controls), for example, use of oral rehydration solutions, zinc, antibiotics, traditional medicines, continued feeding and fluid administration; healthcare seeking

behavior; and breastfeeding practices. The household's direct (out-of-pocket) and indirect (eg, income lost while caring for the sick child) expenditures at home and at the SHC were tabulated. The GEMS staff measured the child's axillary digital temperature, respiratory rate (the average of 2 measures obtained using a rate counter), anthropometric dimensions (described below), and clinical signs of malnutrition (bipedal edema, wasting, flaky skin, and sparse or loose hair).

A clinician examined all cases to document signs of dehydration, including skin pinch return (graded as slow ≤2 seconds or very slow >2 seconds), sunken eyes (more than usual confirmed by the parent/primary caretaker), dry mouth (graded as somewhat or very dry), and mental status changes, and examined the child's rectum for signs of prolapse. A member of the clinical team examined the child's stool (if available) for visible blood and recorded any rehydration fluids, zinc, and antibiotics prescribed or administered at the SHC. Cases who remained in the SHC while receiving rehydration fluids were reweighed at 4 hours and again at discharge from the SHC, as applicable, at which points the clinician reassessed the child for signs of dehydration and determined his/her vital status and weight.

Collection and Processing of Stool Specimens

To qualify for enrollment, each case and control had to produce a whole stool specimen that weighed at least 3 grams. In one site (Kolkata), stool is routinely collected from hospitalized children by passing a small catheter into the child's rectum and aspirating loose stool using a syringe attached to the other end [28]; at all other sites, whole stool was passed naturally per rectum. Cases were required to provide a whole stool specimen within 12 hours of registration at the center. To collect stool from control children at home, study staff provided the caretaker with a polystyrene foam container containing a cold pack, a culturally accepted stool collection device (such as a plastic "potty"), plastic gloves, a specimen cup, and a scoop. The field worker returned to the household the next morning (or sooner if called by the parent) to retrieve the stool sample and perform the study interview. Because children usually defecate in the morning, and since families frequently used cellular phones to alert the GEMS field team that the child had produced a stool, we were able to fulfill the study requirement of retrieving and processing freshly passed stools from cases and controls within 6 hours of evacuation. Processing involved inserting 2 cotton-tipped swabs into the specimen (if dysentery was present, an area of blood or mucus was swabbed); one swab was placed into modified Cary-Blair transport medium [29], and the other into buffered glycerol saline [30]. Remaining whole stool was retained in an empty vial. The processed sample was placed immediately into either a specimen refrigerator or a polystyrene foam container containing a fresh cold pack, to be delivered to the laboratory and plated within 24 hours. Stools were evaluated for bacterial pathogens (eg, *Salmonella*, *Shigella*, *Campylobacter*, *Aeromonas*, and *Vibrio* species, and 5 diarrheal pathotypes of *Escherichia coli*), protozoal agents (*Entamoeba histolytica*, *Giardia lamblia*, and *Crytosporidium* species), and viruses (rotavirus, adenovirus, norovirus, sapovirus, and astrovirus) using microbiologic methods described elsewhere in this supplement (see Panchalingam et al, this supplement).

If antibiotics were to be administered to cases before the whole stool specimen was collected, 2 rectal swabs also were obtained. The cotton tip was moistened with transport media, gently inserted into the child's rectum, rotated 360°, and immediately inserted into transport media, as described above for whole stool swabs. Only swabs stained or covered with fecal material were accepted by the laboratory. This strategy permitted collection of an adequate sample for bacteriology (ie, rectal swabs) prior to antibiotic administration as well as a whole stool for identification of pathogens that are best detected in whole stool but are not expected to be affected by antibiotic administration (see Panchalingam et al, this supplement).

Memory Aid for Recording Diarrheal Episodes in Cases and Controls During the 14 Days After Enrollment

We created a memory aid suitable for use by adults regardless of literacy (Figure 2). The data will be used to detect the occurrence of persistent diarrhea in cases and to explore whether the inclusion of control children who developed diarrhea within 7 days after enrollment impacted the association between specific pathogens and MSD. The tool was developed in collaboration with a representative from the Malian Office of Literacy and modified in response to focus groups and field testing at each site. After receiving training at the enrollment visit, each day for the next 14 days the parent/primary caretakers marked whether the child had normal stools only, or diarrhea (passage of ≥3 abnormally loose stools in the previous 24 hours). The aid was reviewed with the caretaker at the 60-day follow-up visit to resolve missing or unclear markings and then collected. Diarrhea that continued unabated through day 14 will be termed "persistent diarrhea"; diarrheal episodes that continued beyond day 14 (the last day the memory tool collected data) were not systematically tracked.

Clinical and Epidemiologic Data Collected at the Single Household Follow-up Visit

GEMS field workers visited the household of each case and control child approximately 60 days after enrollment (acceptable range, 50–90 days). They assessed vital status, recorded interim medical events, took the child's axillary temperature, and performed anthropometric measurements. They directly observed the household's drinking water sources, storage containers, and treatment practices, and tested the water for

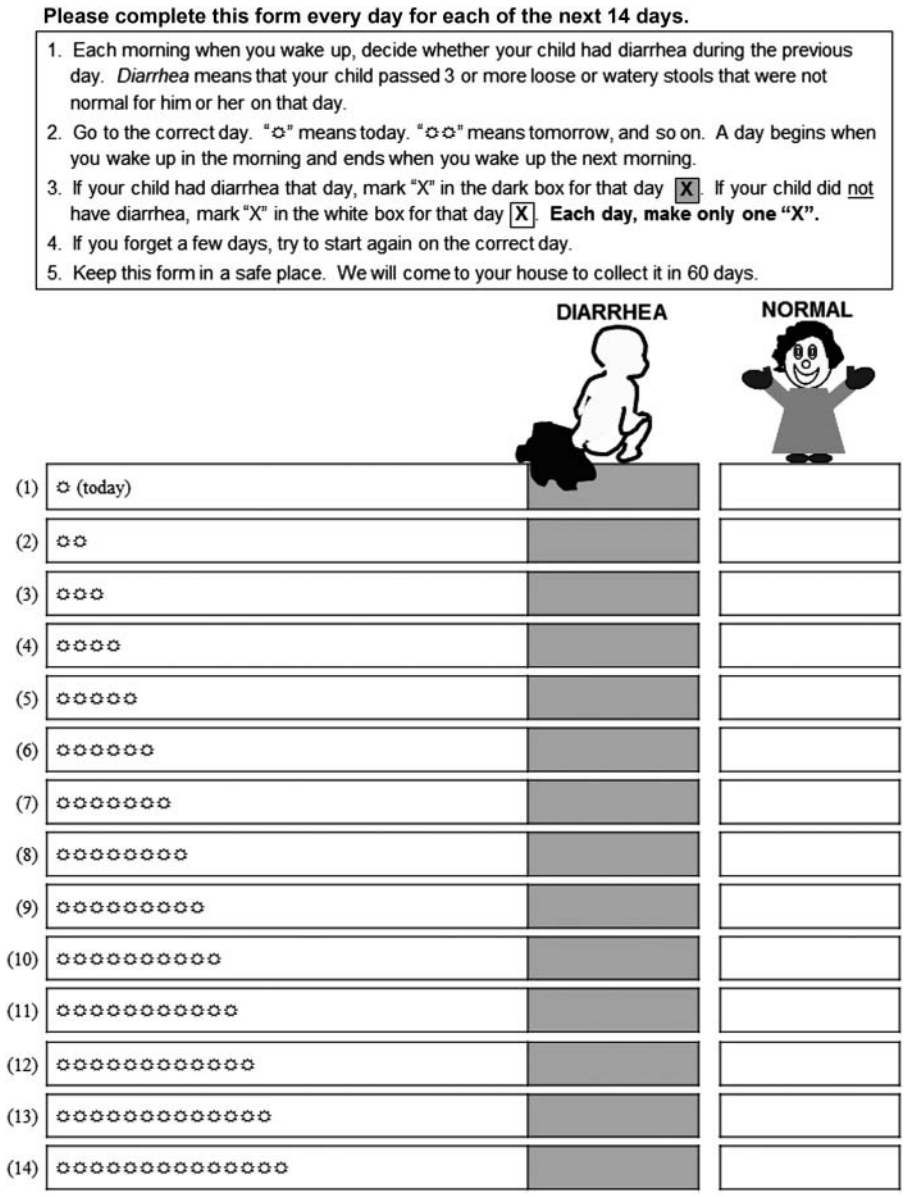

Figure 2. Memory aid completed by the caretaker to document the occurrence of diarrhea for 14 days after enrollment of cases and controls.

chlorine if the household reported that they treated it. They examined the sanitation facilities and noted whether fecal contamination was present, and observed hygiene indicators, such as the proximity of soap to the hand washing station.

Anthropometric Measurements

Weight, length/height, and MUAC were measured for each case and control at enrollment and at the 60-day follow-up visit as previously described [31]. Weight (to the nearest 0.1 kg) was recorded prior to administration of rehydration fluids with the child naked or in light clothing using a digital scale that was calibrated at least weekly (model 314, Tanita Corp of America, Arlington Heights, Illinois); for children 0–23 months of age, the weight of the mother alone and with the child was recorded, and the child's weight was computed during analysis. The length of children 0–23 months of age or those who were older but unable to stand unassisted was measured (to the nearest 0.1 cm) in the recumbent position using a board with a fixed head and sliding foot piece (Shorr Productions, Olney Maryland). The same apparatus was used to measure standing height in children 2 years of age and older. A 25-cm paper single-slotted insertion tape was used to

measure MUAC to the nearest 0.1 cm (Shorr Productions). Length/height and MUAC were each measured thrice; the average will be calculated during analysis [32].

HIV Substudy

In appreciation of the importance of HIV infection on the incidence and outcomes from diarrheal disease, including an increased likelihood of dying from an episode of diarrhea compared with HIV-infected children [33, 34], we considered including systematic HIV testing as part of the initial study design but concluded that it was beyond the scope of our capabilities. As the study progressed, national guidelines for provider-initiated counseling and testing were adopted at the Kenya and Mozambique study sites (the only 2 GEMS study sites with high HIV seroprevalence in adults), and home-based counseling and testing has been implemented at the Kenya site. As a result, during the last 2 years of GEMS, we incorporated voluntary HIV testing or the ability to link to existing HIV test results of mothers and children into the study protocol at these 2 sites. Informed consent was obtained to link HIV test results (for participating child and his/her mother) to GEMS data. We will compare frequency, outcomes, and etiologies of episodes of diarrheal diseases among infected and uninfected children born to infected mothers, and among uninfected children born to uninfected mothers.

Detection of Deaths and Performance of Verbal Autopsy

Two parallel systems were in place at all sites to detect deaths. The GEMS team ascertained deaths among children enrolled in the case/control study during the enrollment encounter and at the 60-day follow-up visit. Concomitantly, the DSS teams identified all under-5 deaths regardless of enrollment status. In either case, the DSS team obtained a verbal autopsy using WHO standardized questionnaires with minor modifications [35]. Local customs were followed to respect the mourning period after which a family could be contacted. Whenever possible, information on the cause of death was collected from the medical chart, the healthcare provider, and the death certificate for use as a means to validate the results of the verbal autopsy [36]. A uniform algorithm will be used across all sites to determine the cause of death. The mortality associated with diarrhea and dysentery among enrolled and nonenrolled children will be calculated.

Sample Size Considerations

A sample size at each site of approximately 600 analyzable cases and 1–3 matched controls per stratum was chosen to provide 80% power (2-sided $\alpha = .05$) for site stratum–specific comparison of the proportion of cases and controls in whom a specific enteropathogen is identified, if a specific pathogen is identified in at least 5.8% of cases and 2.5% of controls. In the event that the proportion of cases with a specific pathogen exceeds 5.8%, the absolute difference between cases and controls needed to achieve statistical significance increases. For example, this sample size will give 80% power to find a significant difference if the proportion is 9.8% in cases vs 5.5% or less in controls. To compensate for dropout, migration, and other losses to follow-up of up to 10%, we planned to enroll a total of 660 cases per stratum per site to achieve the desired number of analyzable cases and controls.

Ethical Considerations and Oversight

GEMS was designed as an observational study that confers minimal risk, and is expected to generate information that can be used by the scientific, public health, policy, and healthcare provider communities to improve the prevention and treatment of diarrheal diseases in the future, both at the site level and globally. Each site was expected to follow WHO guidelines for the clinical diagnosis and management of diarrheal disease, which represent a universal standard of care [18]. We provided supplementary funding for procurement of medical supplies (eg, oral rehydration solution, antibiotics, intravenous cannulae, and fluids) to be used to treat patients with diarrhea at the discretion of each participating SHC. We did not attempt to systematically introduce newer aspects of diarrhea management, such as low-osmolality oral rehydration solution and zinc, in sites lacking national policies to guide usage and trained health providers to administer these products [37]. At the time of study initiation, no site routinely performed the spectrum of assays provided by GEMS to detect potentially treatable pathogens (bacterial culture and immunoassays for protozoa). Therefore, the sites were expected to ensure timely provision of the GEMS results to the clinicians for use in case management. We shared interim results (eg, distribution of pathogens, management of diarrhea, and sequelae) with the investigators and the international community annually at investigators' meetings and at scientific conferences, and each site received a cleaned dataset each year that could be used for more detailed exploration.

The clinical protocol, consent forms, CRFs, and other supporting documents were approved prior to initiation of the study by the ethics committees and applicable scientific review boards at the University of Maryland School of Medicine, and the committees overseeing each site and their collaborating partners from other institutions. Amendments and annual reports underwent ethics committee review. Consent forms were translated into 11 local languages and modified according to the standards at each site; full approval of the University of Maryland ethics committee required back-translation into English and certification by an independent bilingual speaker that the 2 versions were identical. Individual, informed consent was obtained from the parent/primary caretaker of

each participant prior to study activities. When the person supplying consent was illiterate, an impartial third party witnessed the consent process and signed the consent document. Some sites additionally obtained "community consent" from local leaders who were convened in a public forum to discuss the study aims, procedures, potential risks, and benefits.

Data Flow, Management, and Analysis

A data coordinating center (DCC) was responsible for centralized data management as described elsewhere (see Biswas et al, this supplement). In brief, sites transmitted completed CRF pages to the DCC using a variety of electronic formats, but primarily by secure file transfer protocol (SFTP). Although the CRFs were printed in different languages, the structure of the fields was maintained to permit generation of a single database containing data from all sites. The DataFax software system (Clinical DataFax Systems, Hamilton, Ontario, Canada) was used to build and manage the master database and aided in the electronic validation process based on character recognition software. Timelines for transmission of data, data queries, and query resolution were established. A system of security measures and backup procedures preserved the integrity of the data and ensured restoration capabilities.

The GEMS analytic plan (see Blackwelder et al, this supplement) addressed 3 main goals: (1) to determine the major pathogens responsible for MSD, taking into account the prevalence of each pathogen, the frequency of asymptomatic infection in controls, and the presence of multiple pathogens; (2) to determine the pathogen-specific attributable fraction of MSD by age within each site and to extend these estimates to the DSS population; and (3) to identify independent risk factors, including demographic, environmental, and socioeconomic factors as well as pathogens, for MSD and other outcomes of interest (especially death and child growth) using multivariable models.

Quality Management

Activities to ensure high-quality data collection included in-depth training followed by assessment of competency using a variety of techniques, such as written tests and observations (with feedback) during training sessions. To control the quality of data entry, a field supervisor at each site reviewed all completed CRFs daily for legibility, completeness, and consistency. The supervisor's signature indicating that all discrepancies were resolved was required for submission to the DCC. Quality control at the DCC to detect missing data, missing forms, out-of-range values, and data inconsistencies is described elsewhere (see Biswas et al, this supplement).

Supervision and oversight were maintained for quality assurance purposes. Supervisors utilized growth charts and predefined criteria to identify (in real time) aberrant measurements that should be repeated, such values that were lower at 60 days compared to enrollment. Clinical and field supervisors performed random reinterviews to ensure the validity of the data collected by the GEMS clinicians and field workers. The CVD epidemiology team visited each site at least twice per year to observe study activities, review the regulatory files, randomly inspect consent forms and CRFs, and retrain as necessary. They provided a written feedback to the site and the CVD investigators. They maintained at least weekly contact with the teams with the use of email, internet calls, and teleconferences. A regulatory affairs specialist at the CVD oversaw the quality, timeliness, and completeness of submissions to each relevant institutional review board, and ensured that each site was compliant with US regulatory requirements. Sites reported all protocol deviations to the CVD team and a corrective action plan was developed jointly.

DISCUSSION

GEMS is the largest and most comprehensive case/control study of acute diarrhea conducted to date, and will, with its complementary components (HUAS-lite and DSS), provide information about the incidence, microbiologic etiology, risk factors, and adverse clinical outcomes of moderate-to-severe diarrheal episodes among infants and young children living in regions of the world where 82% of under-5 deaths occur [2]. GEMS employed standardized data collection instruments and epidemiologic methods across diverse developing country settings that vary with respect to health indicators, access to and quality of affordable healthcare, economic development, and environmental conditions, so the results will be broadly applicable and can be used to augment existing disease burden models and to define the factors likely to influence the outcome of diarrheal disease in the future. The generalizability of the results should be further enhanced by employing case definitions and study methods that were accepted by experts in the field and disclosed in a detailed and transparent way. GEMS will provide a detailed characterization of MSD according to its clinical manifestations and adverse effects on child health. The attributable fraction contributed by each pathogen that is significantly associated with MSD in the case/control study then will be quantified. This process will produce a list of enteropathogens that should be prioritized for public health interventions. In addition, GEMS will provide the serologic, antigenic, and genotypic characteristics of the major etiologic agents, information needed to develop vaccines and other interventions that can be used in decades to come. We have laid the groundwork for building cost-effectiveness models to justify the introduction of selected interventions into countries where GEMS was undertaken by describing the economic burden of diarrheal diseases.

By conducting the GEMS case/control study within a demographically characterized population whose healthcare utilization practices have been characterized, we are also able to derive population-based estimates of MSD incidence and other outcomes of interest. Repeated HUAS-lite interviews of a representative sample of caretakers over a 2-year period provide the proportion of children in the DSS population who seek care at the SHC when they develop MSD (the *r* value). Because we are measuring the number of MSD cases from the DSS who visit the SHC each year, we can divide by *r* to estimate the number of MSD cases in the DSS population as a measure of incidence. Furthermore, we can use odds ratios from the GEMS case/control study to calculate population attributable fractions for each pathogen found to be associated with MSD by conditional logistic regression adjusted for the presence of other pathogens (see Blackwelder et al, this supplement).

Some limitations of GEMS are noteworthy. First, although designed as an observational study, the resources, infrastructure, training, and frequent visits to the DSS households provided by GEMS may well have altered the natural history of diarrheal diseases in our study populations. These secondary health benefits must be considered in interpreting the outcome and sequelae of diarrheal diseases that are found in GEMS. Another limitation relates to our ability to use the case/control study to derive population-based estimates of MSD incidence and other outcomes of interest. The validity of these estimates will depend on 2 factors. One is the ability of caretakers to accurately report that their child had findings indicative of MSD during the previous 14 days. We attempted to quantify this limitation by conducting a nested study to compare the caretakers' determination of MSD with those of the SHC clinician and found good agreement (κ statistic for interobserver agreement = 0.82 for sunken eyes and 0.64 for wrinkled skin, data not shown). The second factor is the need for a high *r* value to achieve precise estimates and to increase the likelihood that the children enrolled in the case/control study indeed represent those with MSD in the DSS population. For this reason, we endeavored to select the SHCs most likely to capture diarrheal diseases at each site; however, despite our efforts, it was generally not possible to achieve high *r* values. During analysis, we will take into account the variance of the estimated *r*'s in assessing the precision of our disease incidence estimates. Finally, some might argue that our case definition of MSD could bias our pathogen-specific disease estimates by overestimating the importance of agents that cause dysentery. Our decision to include dysentery is supported by published observations indicating that children with dysentery have an increased risk for persistent diarrhea, growth faltering, and death [14, 38, 39]. These concerns can be addressed in the analysis because we know the proportion of all children seeking care at each SHC with MSD that have dysentery, so adjustments can be made as necessary if children with and without dysentery are not equally sampled. We also know the prevalence of dysentery among children in the DSS population based on serial HUAS-lite rounds and can adjust our population-based estimates accordingly.

An important contribution of GEMS is the single follow-up visit to the homes of both cases and controls approximately 60 days after enrollment, which will allow us to elucidate the outcome of children during the vulnerable period that follows acute MSD. The few cross-sectional health center–based studies in developing countries that have contacted children after discharge suggest that most sequelae are missed if surveillance is limited to the hospital stay [40, 41]. A caveat is that we will be unable to prospectively define interim events that might influence the outcomes, including recurring episodes of diarrhea. Important information will be gleaned nonetheless by evaluating the same outcomes in case and control children and determining the relative risk of adverse clinical outcomes during the 50–90 days following an episode of MSD. The lack of interim contact with participants will also impact the quality of data generated by the memory aid regarding the occurrence of diarrhea during the 14 days after enrollment.

In sum, we have described the design and methods of GEMS and our efforts to achieve scientific rigor while maintaining simplicity and standardization. We presented a candid portrait of the considerations that were entertained in developing the study design, the challenges encountered, and solutions developed along with the potential strengths and limitations of the methods. This level of detail is intended to provide the scientific and public health communities with high-quality data that can be used to update and strengthen diarrheal disease burden models and to guide strategic planning and resource allocation for the future.

Notes

Acknowledgments. We are grateful to those who helped us design and prepare for GEMS, especially the families who participated in the training sessions and focus groups and the investigators and staff at the GEMS sites for their enthusiasm and diligence in mastering the study protocol and procedures, including Ciara O'Reilly, Richard Omore, Inacio Mandomando, Tacilta Nhampossa, Sozinho Acacio, Doh Sanogo, Suman Kumar Das, Sujit Bhattacharya, Byomkesh Manna, Philip C. Hill, Adebayo Akinsola, Mitchell Adeyemi, and Farheen Quadri. We thank Paul Stolley for advice in case/control study design, and Karen Ball for support in regulatory affairs. We thank Jan Agosti, Niranjan Bose, Thomas Brewer, and Regina Rabinovich from the Bill & Melinda Gates Foundation for technical support and guidance; Irwin Shorr for providing training in anthropometry; and Cristina Carty, Rebecca Horney, and Barbara Yndo for assistance in developing the data collection and management tools.

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. United Nations Inter-agency Group for Child Mortality Estimation. Levels and Trends in Child Mortality Report 2011. Available at: http://www.unicef.org/media/files/Child_Mortality_Report_2011_Final.pdf. Accessed 16 March 2012.
2. You D, Wardlaw T, Salama P, Jones G. Levels and trends in under-5 mortality, 1990–2008. Lancet **2010**; 375:100–3.
3. Walker CL, Aryee MJ, Boschi-Pinto C, Black RE. Estimating diarrhea mortality among young children in low and middle income countries. PLoS One **2012**; 7:e29151.
4. Anker M. The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. Int J Epidemiol **1997**; 26:1090–6.
5. Lozano R, Lopez AD, Atkinson C, Naghavi M, Flaxman AD, Murray CJ. Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards. Popul Health Metr **2011**; 9:32.
6. Morris SS, Black RE, Tomaskovic L. Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems. Int J Epidemiol **2003**; 32:1041–51.
7. Lawn JE, Wilczynska-Ketende K, Cousens SN. Estimating the causes of 4 million neonatal deaths in the year 2000. Int J Epidemiol **2006**; 35:706–18.
8. Cooper RS, Osotimehin B, Kaufman JS, Forrester T. Disease burden in sub-Saharan Africa: what should we conclude in the absence of data? Lancet **1998**; 351:208–10.
9. World Health Organization Multicentre Growth Reference Study Group. WHO child growth standards. Methods and development: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. Available at: http://www.who.int/childgrowth/standards/technical_report/en/index.html. Accessed 20 May 2012.
10. World Health Organization. Handbook for good clinical research practice (GCP): guidance for Implementation. Available at: http://whqlibdoc.who.int/publications/2005/924159392X_eng.pdf. Accessed 21 May 2012.
11. World Health Organization Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl **2006**; 450:38–46.
12. Baqui AH, Black RE, Yunus M, Hoque AR, Chowdhury HR, Sack RB. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. Int J Epidemiol **1991**; 20:1057–63.
13. Schmidt WP, Arnold BF, Boisson S, et al. Epidemiological methods in diarrhoea studies—an update. Int J Epidemiol **2011**; 40:1678–92.
14. Black RE, Brown KH, Becker S, Abdul-Alim ARM, Huq I. Longitudinal studies of infectious diseases and physical growth in rural Bangladesh. II. Incidence of diarrhea and association with known pathogens. Am J Epidemiol **1982**; 115:315–24.
15. Black RE, Brown KH, Becker S. Effects of diarrhea associated with specific enteropathogens on the growth of children in rural Bangladesh. Pediatrics **1984**; 73:799–805.
16. Butler T, Islam M, Azad AK, Islam MR, Speelman P. Causes of death in diarrhoeal diseases after rehydration therapy: an autopsy study of 140 patients in Bangladesh. Bull World Health Organ **1987**; 65:317–23.
17. Dutta P, Mitra U, Rasaily R, et al. Assessing the cause of in-patients pediatric diarrheal deaths: an analysis of hospital records. Indian Pediatr **1995**; 32:313–21.
18. World Health Organization. Handbook: IMCI integrated management of childhood illness. Available at: http://www.who.int/maternal_child_adolescent/documents/9241546441/en/index.html. Accessed 22 March 2012.
19. UNICEF/World Health Organization. Diarrhoea: why children are still dying and what can be done. Available at: http://www.who.int/topics/diarrhoea/en/. Accessed 22 February 2012.
20. Ruuska T, Vesikari T. Rotavirus disease in Finnish children: use of numerical scores for clinical severity of diarrhoeal episodes. Scand J Infect Dis **1990**; 22:259–67.
21. Flores J, Perez-Schael I, Gonzalez M, et al. Protection against severe rotavirus diarrhoea by rhesus rotavirus vaccine in Venezuelan infants. Lancet **1987**; 1:882–4.
22. 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**; 158:570–87.
23. Ferreccio C, Prado V, Ojeda A, et al. Epidemiologic patterns of acute diarrhea and endemic *Shigella* infections in a poor periurban setting in Santiago, Chile. Am J Epidemiol **1991**; 134:614–27.
24. Levine MM, Ferreccio C, Prado V, et al. Epidemiologic studies of *Escherichia coli* infections in a low socioeconomic level periurban community in Santiago, Chile. Am J Epidemiol **1993**; 138: 849–69.
25. Wright JA, Gundry SW, Conroy R, et al. Defining episodes of diarrhoea: results from a three-country study in sub-Saharan Africa. J Health Popul Nutr **2006**; 24:8–16.
26. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography **2001**; 38:115–32.
27. World Health Organization/UNICEF. WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation Definition and Methods. Available at: http://www.wssinfo.org/. Accessed 6 March 2012.
28. Wang XY, Ansaruzzaman M, Vaz R, et al. Field evaluation of a rapid immunochromatographic dipstick test for the diagnosis of cholera in a high-risk population. BMC Infect Dis **2006**; 6:17.
29. Cary SG, Blair EB. New transport medium for shipment of clinical specimens. I. Fecal specimens. J Bacteriol **1964**; 88:96–8.
30. Wells JG, Morris GK. Evaluation of transport methods for isolating *Shigella* spp. J Clin Microbiol **1981**; 13:789–90.
31. Cogill B. Anthropometric indicators measurement guide. Available at: http://www.fantaproject.org. Accessed 19 August 2007.
32. World Health Organization. WHO Anthro for personal computer, version 3, 2009. Available at: http://www.who.int/childgrowth/software/en/. Accessed 22 February 2010.
33. Thea DM, St Louis ME, Atido U, et al. A prospective study of diarrhea and HIV-1 infection among 429 Zairian infants. N Engl J Med **1993**; 329:1696–702.
34. Madhi SA, Cunliffe NA, Steele D, et al. Effect of human rotavirus vaccine on severe diarrhea in African infants. N Engl J Med **2010**; 362:289–98.
35. World Health Organization. Verbal autopsy standards: ascertaining and attributing cause of death. Available at: http://www.who.int/healthinfo/statistics/verbalautopsystandards/en/. Accessed 28 March 2012.
36. Anker M, Black RE, Coldham C, et al. A standard verbal autopsy method for investigating causes of death in infants and children (WHO document WHO/CDS/CSR/ISR/99.4). Available at: http://www.who.int/csr/resources/publications/surveillance/WHO_CDS_CSR_ISR_99_4/en/. Accessed 22 March 2012.

37. World Health Organization, United Nations Children's Fund, US Agency for International Development, and Johns Hopkins Bloomberg School of Public Health. Implementing the new recommendations on the clinical management of diarrhea: guidelines for policy makers and programme managers. Available at: http://www.zinctaskforce.org/policy-advocacy/. Accessed 22 March 2012.
38. Ahmed F, Ansaruzzaman M, Haque E, Rao MR, Clemens JD. Epidemiology of postshigellosis persistent diarrhea in young children. Pediatr Infect Dis J **2001**; 20:525–30.
39. Bennish ML, Wojtyniak BJ. Mortality due to shigellosis: community and hospital data. Rev Infect Dis **1991**; 13(suppl 4):S245–51.
40. Greenwood BM, Greenwood AM, Bradley AK, Tulloch S, Hayes R, Oldfield FS. Deaths in infancy and early childhood in a well-vaccinated, rural, West African population. Ann Trop Paediatr **1987**; 7:91–9.
41. Stanton B, Clemens J, Khair T, Shahid NS. Follow-up of children discharged from hospital after treatment for diarrhoea in urban Bangladesh. Trop Geogr Med **1986**; 38:113–8.
42. United Nations Children's Fund (UNICEF). The state of the world's children 2005. Available at: http://www.unicef.org/sowc05/english/fullreport.html. Accessed 29 March 2012.
43. World Health Organization. WHO world malaria report 2005. Available at: http://rbm.who.int/wmr2005/. Accessed 19 August 2009.

SUPPLEMENT ARTICLE

Statistical Methods in the Global Enteric Multicenter Study (GEMS)

William C. Blackwelder,[1] Kousick Biswas,[2] Yukun Wu,[1] Karen L. Kotloff,[1] Tamer H. Farag,[1] Dilruba Nasrin,[1] Barry I. Graubard,[3] Halvor Sommerfelt,[4,5] and Myron M. Levine[1]

[1]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, [2]Veterans Administration, Perry Point, and [3]National Cancer Institute, Bethesda, Maryland; [4]Centre for International Health, University of Bergen, and [5]Division of Infectious Disease Control, Norwegian Institute of Public Health, Oslo, Norway

The Global Enteric Multicenter Study (GEMS) is an investigation of the burden (number of cases and incidence) of moderate-to-severe diarrhea (MSD) in children <60 months of age at 7 sites in sub-Saharan Africa and South Asia. The population attributable fraction for a putative pathogen, either unadjusted or adjusted for other pathogens, is estimated using the proportion of MSD cases from whom the pathogen was isolated and the odds ratio for MSD and the pathogen from conditional logistic regression modeling. The adjusted attributable fraction, proportion of MSD cases taken to a sentinel health center (SHC), number of cases presenting to an SHC, and the site's population are used to estimate the annual number of MSD cases and MSD incidence rate attributable to a pathogen or group of pathogens. Associations with death and nutritional outcomes, ascertained at follow-up visits to case and control households, are evaluated both in MSD cases and in the population.

Diarrheal diseases are one of the top 2 causes of death among children <60 months of age in the developing world [1]. Interventions to diminish this enteric disease burden among the world's most disadvantaged pediatric populations are expected to include pathogen-specific vaccines and diagnostics (followed by specific treatment), as well as expanded use of nonspecific therapeutic regimens such as oral rehydration and zinc. Despite a plethora of individual site studies [2–8], and a few coordinated multicountry studies [9] of both case/control and prospective cohort design, there remains much disagreement over the relative importance of various specific bacterial, viral, and protozoal pathogens as causes of diarrheal illness, particularly of clinically more severe forms. As additional diarrheal pathogens have come to be described in recent years and as diagnostic microbiologic tests have become more sensitive, the need for a definitive study of pediatric diarrheal disease gained widespread support and momentum [10]. This led to initiation of the Global Enteric Multicenter Study (GEMS), a matched case/control study of the burden, in terms of numbers of cases and incidence rates, of moderate-to-severe diarrhea (MSD) in children <60 months of age at 4 sites in sub-Saharan Africa and 3 sites in South Asia (see Levine et al and Kotloff et al in this supplement) [11, 12]. GEMS, which involves the detection of a wide array of etiologic agents in enrolled MSD cases and their matched controls, represents a historic multisite undertaking to apply standardized specific microbiologic methods to detect evidence of infection with 1 or more of a wide array of potential pathogens, and to use the resulting data to estimate the disease burden attributable to specific pathogens.

In this paper we describe and illustrate the major statistical methods used in GEMS. These include methods for assessing associations between the presence of specific pathogens (and other variables of

Correspondence: William C. Blackwelder, PhD, Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201 (wblackwe@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S246–53

DOI: 10.1093/cid/cis788

interest) and MSD, estimating the proportion and absolute number of cases of MSD due to specific pathogens, estimating the incidence of MSD and the proportion of MSD cases who are taken to one of the health centers designated as a study site (referred to as sentinel health centers [SHCs]), and assessing associations between pathogens and other variables of interest with outcomes other than MSD.

STUDY DESIGN

For GEMS, MSD is defined by an episode of diarrhea (≥3 loose stools within a 24-hour period) with onset within the past 7 days and at least 7 days after the end of any previous episode, and at least 1 of the following: sunken eyes, more than normal; loss of skin turgor; intravenous rehydration administered or prescribed; visible blood in stool; or hospitalization with diarrhea. At each GEMS study site, children 0–59 months of age who had MSD and were brought to one of the site's SHCs were enrolled into the study, along with 1 or more controls who were matched to the case by age, time (within 7 days of the case enrollment), and geographic location of residence. Up to 8 or 9 cases per 2-week period in each of 3 age strata (0–11, 12–23, and 24–59 months) were typically enrolled during a 3-year enrollment period at each site. These strata represent age groups in which MSD and its clinical presentations (eg, dysentery) are observed with different frequencies and when the etiologies are known to be somewhat different. For example, certain etiologies are relatively more important in infants, while others are more common in toddlers or preschool children with MSD. Most of our analyses have been done within these age strata. Calculation of statistical power was based on comparing 2 independent proportions; for a moderate degree of correlation between presence of a pathogen in a case and in its matched control, the power for a given sample size will be higher for a test designed for matched data. The planned sample size at each site was 600 analyzable case/control pairs. This sample size should be sufficient, for example, for a test at the 2-sided 5% significance level to have 80% power to find a significant difference between proportions of cases and controls for which a specific pathogen is isolated, if the respective true proportions are 5.8% and 2.5%. Stool specimens were collected from each case and control for identification of potential enteric pathogens. Demographic, anthropometric, and other information about the study child and the household was collected at enrollment and also at a home visit approximately 50–90 days after enrollment.

Besides the matched case/control study, a Health Care Utilization and Attitudes Survey (HUAS), based on random sampling from each site's demographic surveillance system (DSS), was conducted before the beginning of the study. Truncated versions of the HUAS, known as HUAS-lite, were conducted several times during the 3 years of case and control enrollment. The HUAS was used to evaluate associations between a variety of demographic factors and characteristics of households (eg, main source of water, main method of disposal of feces) with the presence of diarrhea in the sampled child. The HUAS-lite surveys are used primarily to estimate the proportion of children with MSD who were taken to one of the site's SHCs and to estimate the 1-week incidence of MSD.

The GEMS study design is given in more detail in the article by Kotloff et al in this supplement [12].

ANALYSES OF HUAS AND HUAS-LITE DATA

The primary analyses of the HUAS and HUAS-lite data are (1) estimation of the proportion of children with MSD who are taken to one of the site's SHCs within 7 days of onset of diarrhea, (2) estimation of the 1-week incidence of MSD, and (3) identification of associations between characteristics of a household or primary caretaker and care seeking for diarrhea.

We use "r" to represent the proportion (and its estimate) of MSD cases who were taken to one of our designated SHCs within 7 days of onset of diarrhea; r is calculated from HUAS-lite data, since the HUAS-lite surveys were conducted during the period of case/control enrollment. In calculating r for all sites except Kenya, we use site-specific sampling weights that are defined for each combination of age group (0–11, 12–23, and 24–59 months) and sex. (The entire DSS population is included in HUAS-lite surveys in the Kenya site, so no weighting is necessary.) For each HUAS-lite round, the sampling weight for an age-sex category is the number of children in that category in the DSS population represented by each child in the HUAS-lite sample. Then the weight for each child in an age-sex category is the DSS population total for that category divided by the number of children in the HUAS-lite sample in the category. These weights are used in a time-to-event (life table) analysis using the Kaplan-Meier method to estimate, at each day beginning with the day after onset of diarrhea, the proportion of children with MSD in the population who had been taken to an SHC. Time-to-event analysis is used because for many children who currently had MSD, the HUAS-lite interview was conducted before the child reached the seventh day of the episode, so the child had not had a full 7 days after onset of diarrhea in which to be taken to an SHC. The data for all HUAS-lite rounds conducted during the case/control study are pooled [13], with each child weighted according to the sampling weight assigned for that child's HUAS-lite round. The estimate of r is then the proportion of children who were taken to an SHC by day 7 after onset of diarrhea in the time-to-event analysis.

We also estimate the 1-week incidence of MSD from HUAS-lite data. For this estimate, we pool data from different HUAS-lite rounds and use sampling weights to obtain these estimates, as is done for estimates of the proportion of cases taken to an SHC within 7 days of onset. We count the number of children with MSD whose illness began on the day of the interview or one of the 6 days preceding that day. As for estimating r, time-to-event analysis is used. In this analysis, children with diarrhea that had not progressed to MSD and whose diarrhea began <1 week prior to the HUAS-lite survey were censored after the number of days they had had diarrhea.

Because much more information was collected in the HUAS than in the HUAS-lite rounds, with the HUAS data we study associations between care seeking for diarrhea and a variety of characteristics of a household or primary caretaker. The main analytic method used to assess these associations is logistic regression modeling, in which we use the sampling weights in order to obtain results that relate to the DSS population.

ASSOCIATIONS WITH MSD IN THE CASE/CONTROL STUDY

An analysis that is central to the aims of GEMS is the evaluation from the case/control data of potential risk and protective factors for MSD. Because cases of MSD are matched with 1 or more controls, we use conditional logistic regression (CLR) modeling to estimate associations with MSD [14]. In this type of model, case status (case = 1, control = 0) is the dependent variable. The model differs slightly from the usual (unconditional) logistic regression model in that there is no intercept. Thus, the fitted model is of the form

$$\log_e(\text{odds}) = b_1x_1 + b_2x_2 + \cdots + b_kx_k,$$

where $x_1, x_2, \ldots, x_k$ are independent variables under study for association with MSD and $b_1, b_2, \ldots, b_k$ are estimates of the corresponding coefficients. When a variable x_i is dichotomous (ie, 1 if the factor is present and 0 if the factor is not present), $\exp(b_i)$ is an estimate of the odds ratio for the factor—that is, the ratio of the odds of MSD when the factor is present to the odds when the factor is absent, where the odds of an event that occurs with probability Q are Q/(1-Q). In order to obtain appropriate results when the number of discordant case/control pairs (ie, pairs where the factor is present in the case and absent in the control, or vice versa) is 0, we use a penalized likelihood approach [15]. Typically, we fit CLR models for each site and age category separately. In certain analyses it may be appropriate to combine data for different sites and/or different age groups.

Of special interest are associations of putative enteric pathogens—bacterial, viral, and protozoan—with MSD. To assess and quantify the contribution of a specific pathogen without regard to the presence of other pathogens, we fit a CLR model with a dichotomous variable, representing presence or absence of the pathogen, as the only covariate. In analysis of the contribution of a pathogen adjusted for other specific pathogen (s), we fit models with multiple dichotomous variables, each representing presence or absence of one of the pathogens, as covariates. In developing these models, interactions between the effects of pairs of pathogens are considered (ie, the possibility that the association of a pathogen with MSD depends on whether another pathogen is present).

CLR modeling is also used to evaluate associations of environmental and socioeconomic factors with MSD. Two analyses of particular interest are of associations of water sources, sanitary facilities, and hygiene practices with MSD, and of care-seeking costs with MSD.

THE POPULATION ATTRIBUTABLE FRACTION OF MSD DUE TO 1 OR MORE PATHOGENS

The population attributable fraction (AF) of a disease due to a risk factor is the proportion of disease cases (or the proportion of the risk of disease) that might theoretically be eliminated if the risk factor were eliminated. Other names that have been used for this concept include attributable risk and etiologic fraction. AF can be estimated equivalently [16] from the distribution of the exposure (risk factor) either in the entire population [17] or in cases of disease [18]. Although the concept of AF has been known and applied for decades and there have been scores of case/control and cohort studies that have tested for multiple etiologic agents of diarrheal disease to gather information on the relative importance of different agents in association with diarrhea, we have noted only 2 etiologic studies of diarrhea in which the AF concept was applied [8, 19].

In GEMS we use AF to estimate the fraction of MSD cases due to a specific pathogen or a group of pathogens. We calculate AF for a pathogen, A, as though A were the only risk factor for MSD, and also for A adjusted for other pathogens that might be present. Adjustment for other pathogens is important, since at least 2 of the potential pathogens under study in GEMS were identified in substantial percentages of both cases and controls.

To determine what pathogens are associated with MSD in GEMS, we fit CLR models, as described above, to the data on cases and matched controls. Unadjusted AF for pathogen A is estimated from a model in which the only covariate is an indicator variable y for the presence of A (ie, y = 1 if A is present and y = 0 if A is absent). Given the coefficient b of y in the fitted model, the odds ratio (OR) for MSD and A is estimated

as e^b. We assume we have a random sample of MSD cases and represent the proportion of MSD cases for which A is present by Pr (A|MSD). If OR >1, the unadjusted attributable fraction AF_u is then given by

$$AF_u = \Pr(A|MSD)\left(1 - \frac{1}{OR}\right). \quad (1)$$

As is common in case/control studies, we use OR as an approximation to the risk ratio (RR). In GEMS the 1-week incidence of MSD, which is the basis for choosing cases, is small (ranging from <1% to approximately 9%, depending on the study site and age group). Since the controls are closely matched in time to cases, OR is a close approximation to the incidence rate ratio [20], which with these small incidence rates is in turn close to the RR.

We estimate AF for pathogen A1, adjusted for the presence of other pathogens, as in Bruzzi et al [21]. For example, suppose we adjust A1 for another pathogen A2. We fit a (multiple) conditional logistic regression model that in its most general form includes variables y_1 and y_2, indicating presence or absence of A1 and A2, respectively; and the product of y_1 and y_2, which represents the interaction of the effects of A1 and A2. In this model an interaction indicates that the OR for MSD and A1 depends on whether or not A2 is present. The model will have estimated coefficients b_{10} for y_1, b_{01} for y_2, and b_{11} for the product of y_1 and y_2. We let ρ_{ij} be the proportion of cases with $y_1 = i$ and $y_2 = j$, for i and j = 0 or 1; for example, ρ_{10} is the proportion of cases with A1 present and A2 absent. Then AF_a, the adjusted attributable fraction estimate for A1, can be written

$$AF_a = \rho_{10}\left(1 - \frac{1}{T_{10}}\right) + \rho_{11}\left(1 - \frac{1}{T_{11}}\right), \quad (2)$$

where $T_{10} = \exp(b_{10})$ and $T_{11} = \exp(b_{10} + b_{11})$. Note that the coefficient b_{01} does not appear in formula (2); only coefficients corresponding to the presence of A1 are included. T_{1j} is the ratio of the OR for the combination (1j), in which A1 is present, to the OR for (0j), the same combination of pathogens except that A1 is absent. If there is no interaction term in the model, $AF_a = (\rho_{10} + \rho_{11})\ (1–1/T_{10}) = \Pr\ (A1|MSD)\ (1–1/T_{10})$, where T_{10} is now simply the odds ratio for A1 when A2 is absent; in this case AF_a has the same form as AF_u in equation (1), the only difference being that the OR T_{10} is estimated from a model that includes both y_1 and y_2.

In general, we can estimate the combined attributable fraction for a set of 1 or more pathogens, each with AF >0, possibly adjusted for 1 or more other pathogens—that is, in the above description of AF for A1 adjusted for A2, both pathogens A1 and A2 can be replaced by multiple pathogens. For example, suppose we want estimate to AF_a for 2 pathogens A1 and A2 (set I, the pathogens for which a summary AF is desired), adjusted for pathogens A3 and A4 (set II, the pathogens included only for conditioning). In this case we consider proportions ρ_{ijkl}, representing all combinations of presence or absence of the 4 pathogens; in the GEMS data, some of these proportions will be 0, since specific combinations with more than 2 pathogens occur infrequently. In its most general form, the model will include indicator variables y_1, y_2, y_3, and y_4, and all possible products (interactions) of these variables. The formula for AF_a is

$$AF_a = 1 - \frac{\Sigma\rho_{ijkl}}{T_{ijkl}}. \quad (3)$$

In formula (3), the summation is over all i, j, k, l = 0 or 1; T_{ijkl}, the term for combination (ijkl), is a ratio of the OR for the combination and an OR when the variables representing pathogens in set I (pathogens A1 and A2, for which a summary AF is to be estimated) are all set to 0. T_{ijkl} thus includes coefficients that correspond to pathogens in set I that are present in the combination, as well as interactions between any of them and pathogens in set II (pathogens A3 and A4, the conditioning set) that are present in the combination. T_{ijkl} will not include any coefficients of "main effects" for pathogens in set II or interactions between pathogens in set II, because these appear in both the numerator and denominator ORs that determine T_{ijkl} [21]. For a combination of the form (00kl), all pathogens in set I are absent, and $T_{ijkl} = 1$.

Suppose the CLR model for 4 pathogens, the first 2 in the set for which AF is to be estimated and the last two in the conditioning set, includes terms for all main effects and 2-way interactions (but not higher-order interactions), with coefficients b_{ijkl} for combination (ijkl). Then, for example, the combination of pathogens 1, 2, and 4 corresponds to $T_{1101} = \exp(b_{1000} + b_{0100} + b_{1100} + b_{1001} + b_{0101})$. Note that a coefficient with >1 of the subscripts equal to 1 is the coefficient of an interaction term; b_{1100} is the coefficient of the product y_1y_2, etc. For this example Table 1 gives all possible values of the natural logarithm of T, $\log_e(T_{ijkl})$, in terms of the estimates b_{ijkl} from the CLR model. Note that any combination (ijkl) for which no pathogen from set I is present will have $\log_e(T_{ijkl}) = 0$ (ie, $T_{ijkl} = 1$) in formula (3).

In the GEMS data we have occasionally seen evidence of 2-way interactions, but we have seen no evidence of 3-way or higher interactions; thus, only interactions involving 2 pathogens need be considered in the GEMS analysis.

Cases were sampled for GEMS in approximately equal numbers during each 2-week period, regardless of the number of MSD cases appearing at the SHCs. We estimate AF both unweighted and using weights defined as (number of eligible cases/number of enrolled cases), ie, as the inverse of the

Table 1. Natural Logarithm of Factors (Ratios of Odds Ratios) Corresponding to Combinations of Pathogens in Example of Adjusted Attributable Fraction (AF) Calculation: AF Is Calculated for Pathogens A1 and A2, Adjusted for Pathogens A3 and A4

Pathogen(s) Present	ijkl	$\log_e(T_{ijkl})$
A1	1 0 0 0	b_{1000}
A2	0 1 0 0	b_{0100}
A1, A2	1 1 0 0	$b_{1000} + b_{0100} + b_{1100}$
A1, A3	1 0 1 0	$b_{1000} + b_{1010}$
A2, A3	0 1 1 0	$b_{0100} + b_{0110}$
A1, A2, A3	1 1 1 0	$b_{1000} + b_{0100} + b_{1100} + b_{1010} + b_{0110}$
A1, A4	1 0 0 1	$b_{1000} + b_{1001}$
A2, A4	0 1 0 1	$b_{0100} + b_{0101}$
A1, A2, A4	1 1 0 1	$b_{1000} + b_{0100} + b_{1100} + b_{1001} + b_{0101}$
A1, A3, A4	1 0 1 1	$b_{1000} + b_{1010} + b_{1001}$
A2, A3, A4	0 1 1 1	$b_{0100} + b_{0110} + b_{0101}$
A1, A2, A3, A4	1 1 1 1	$b_{1000} + b_{0100} + b_{1100} + b_{1010} + b_{0110} + b_{1001} + b_{0101}$
A3	0 0 1 0	0
A4	0 0 0 1	0
A3, A4	0 0 1 1	0
None	0 0 0 0	0

sampling fraction for MSD cases. Data for adjacent 2-week periods are combined when there are no enrolled cases in a period. AF estimation is done separately for the 3 age strata (0–11, 12–23, and 24–59 months) within which MSD cases were sampled.

Table 2 shows unadjusted and adjusted AF estimates, unweighted, for children aged 12–23 months in India for the first 2 years of the 3-year GEMS case/control study. These results are typical of the results for other sites and age groups, in that there are few important interactions and the adjusted estimates are not very different from the unadjusted estimates. In particular, with few exceptions, the adjusted AF and number of attributable cases for the major pathogens change only modestly, compared to the unadjusted estimates.

CALCULATION OF ATTRIBUTABLE MSD CASES AND MSD INCIDENCE

Let M_{CC}, M_{SHC}, and M_{pop} represent the total number of MSD cases enrolled in the study, the total numbers of MSD cases seen at the site's SHCs, and the total number of MSD cases in the population in 3 years, respectively. Then for each site and age category, the respective numbers of cases attributable to pathogen A in the study, in the site's SHCs, and in the population are given by $AF \times M_{CC}$, $AF \times M_{SHC}$, and $AF \times M_{pop}$, respectively.

The numbers of cases attributable to A in the study and the SHCs are calculated directly from AF and the numbers of cases, since we observe M_{CC} and take M_{SHC} as the number of cases presenting at the SHCs who are eligible for the study. However, we do not observe M_{pop} directly, but rather estimate it from M_{SHC} and the estimated proportion, r, of MSD cases taken to one of the study site's SHCs. As indicated above, r is estimated from the HUAS-lite rounds conducted during the study.

The estimated annual number of MSD cases in the population during the 3-year case/control study period is $M_{pop} = M_{SHC}/(3r)$, and the estimated annual number of cases attributable to A is $AF \times M_{pop} = AF \times M_{SHC}/(3r)$. If N is the average population at the site over the study period, the annual incidence rate of MSD attributable to A during the study is

Table 2. Crude and Adjusted Attributable Fraction and Attributable Number of Cases in First 2 Years of the Global Enteric Multicenter Study: India, Ages 12–23 Months (364 Cases, 374 Controls)

		Unadjusted Analysis				Adjusted Analysis			
Pathogen	Cases With Pathogen	OR[a]	*P* Value[b]	AF	Attributable Cases	OR[a]	*P* Value[b]	AF	Attributable Cases
Rotavirus	104	22.5	<.0001	0.273	99	36.4	<.0001	0.278	101
Shigella	30	11.4	.0003	0.075	27	38.9	<.0001	0.080	29
ETEC LT/ST or ST	34	2.6	.004	0.057	21	4.3	.0006	0.072	26
Cryptosporidium	45	1.7	.031	0.052	19	2.4	.006	0.073	26
Vibrio cholerae O1	19	8.8	.002	0.046	17	9.1	.002	0.046	17
Adenovirus 40/41	18	6.0	.003	0.041	15	9.5	.002	0.044	16
Entamoeba histolytica	7	3.0	.15	0.013	5	10.3	.038	0.017	6

Abbreviations: AF, attributable fraction; ETEC, enterotoxigenic *Escherichia coli*; LT, heat-labile enterotoxin; OR, odds ratio; ST, heat-stable enterotoxin.

[a] OR: ratio of the odds of moderate-to-severe diarrhea when the putative pathogen is present to the odds when it is absent. OR >1 indicates a positive association.

[b] *P* value from logistic regression.

Table 3. Annual Attributable Moderate-to-Severe Diarrhea Cases in Population and Incidence per 100 Child-Years in First 2 Years of the Global Enteric Multicenter Study: The Gambia, Ages 0–11 Months (312 Cases, 398 Controls)

Pathogen	AF	Annual Attributable Cases[a]	Attributable Incidence Rate[b]
Rotavirus	0.211	135	2.3
Cryptosporidium	0.095	61	1.0

M_{SHC} = No. of eligible MSD cases at SHCs (observed) = 625; r = proportion of MSD cases seen at SHC = 0.487; M_{pop} = annual MSD cases in population = $M_{SHC}/(2r)$ = 642; N = No. of children in population = 5922.

Abbreviations: AF, attributable fraction; MSD, moderate-to-severe diarrhea; SHC, sentinel health center.

[a] Calculated as $AF \times M_{SHC}/(2r)$.

[b] Per 100 child-years; calculated as $100 \times AF \times M_{SHC}/(2rN)$.

approximately $AF \times M_{SHC}/(3rN)$. N is estimated as the median of population estimates from several DSS rounds performed during the study. Table 3 illustrates these calculations for data from the first 2 years of the study on infants aged 0–11 months in The Gambia.

The variance of the incidence rate is approximated by Taylor series to first derivative terms (delta method). The variance of AF is estimated using a jackknife procedure [13], the variance of r as the variance of the probability of an event in a weighted Kaplan-Meier analysis, the variance of $M_{SHC}/3$ as the variance of the mean of 3 yearly totals of cases coming to an SHC, and the variance of N as the variance of the median of several observations from a normal distribution [22].

ASSOCIATIONS WITH OUTCOMES IN CASES OF MSD

Because the GEMS case/control study includes follow-up visits at approximately 60 days (range, 50–90 days), we have information on certain outcomes. Among these are death and, among cases and controls who survive to the follow-up visit, linear and ponderal growth. It is thus natural to investigate risk or protective factors for these outcomes in MSD cases. For a dichotomous outcome such as death, we use logistic regression modeling or, in order to use the actual follow-up times, Cox proportional hazards regression. For a continuous outcome we use linear regression analysis. This analysis is especially relevant for prioritizing the development of point-of-care diagnostics and therapeutic interventions. For a pathogen that is associated with a high case-fatality rate, it might also suggest a need for a prophylactic intervention, such as a vaccine.

ASSOCIATIONS WITH OUTCOMES IN THE POPULATION

Besides evaluating associations between outcomes and risk or protective factors in cases, we are interested in such associations in the general population. This type of analysis is particularly well suited for evaluating the need for preventive interventions, such as vaccines, that target specific pathogens or environmental conditions and would be widely applied in the population. To evaluate this type of association, we use a weighted analysis of cases and controls, which is described in detail by Sommerfelt et al in this supplement [23]. In this approach, weights are chosen so as to make the proportion of cases in the analysis approximately the same as in the population from which the cases were drawn.

DISCUSSION

Of the various statistical analyses in GEMS, it is the analysis of etiology that is the most important and demanding. The reason is that one of the driving rationales for initiating the GEMS was to be able, on the basis of the results, to prioritize the allocation of financial and other resources toward the implementation of existing interventions (such as vaccines and therapeutics) and to prioritize investments in research aimed at developing new interventions, based on the relative contributions of different pathogens to the overall burden of MSD in young children. From this perspective, one sees clearly the potential utility of the AF, defined as the proportion of MSD that would be eliminated if the target population were no longer exposed to a specific risk factor (such as a specific pathogen). AF allows us to estimate the number of MSD cases at one of our sites that can be attributed to a specific pathogen, adjusted for other pathogens that might also be present. Thus, we can distinguish between a pathogen that is responsible for a large number of cases and another pathogen that might be associated with MSD but for which the number of attributable cases is considerably smaller. This is crucial in allowing policy makers to set priorities for interventions. Further, it can help us to identify locations where a specific intervention might make a large impact and other locations where its impact might be relatively minor.

There are, of course, limitations of our study and analysis. The most important limitation regarding statistical analysis is probably in the estimation from HUAS-lite surveys of the proportion of MSD cases taken to an SHC, which we call "r." This proportion is important in our estimation of total MSD burden, as well as the burden attributable to specific pathogens. The HUAS-lite surveys were based on random samples from the DSS population. However, for various reasons (eg, in Kolkata many MSD cases were taken to private healthcare

providers) in most sites and age groups r is smaller than we had hoped it would be (<40% in all except 1 site). Thus, there is the potential for bias in the proportion of cases enrolled with a specific manifestation of MSD or in which a specific pathogen was isolated. A minor limitation is that our requirement that eligible cases should have had onset of diarrhea within the past 7 days could produce a slight underestimate of the true incidence of MSD. Data from the HUAS and HUAS-lite surveys, including estimates of r, will be presented in papers that are in preparation.

Various bacterial, viral, and protozoal pathogens can each cause MSD in children, and some set of these enteropathogens are collectively responsible for most of the MSD that occurs among young children in developing countries. These diarrheal pathogens are transmitted to susceptible children in 1 or more ways, depending on the pathogen, including via contaminated water or food, direct contact with fecally contaminated hands, flies acting as mechanical vectors, and contaminated fomites. There are 2 broad approaches to diminish MSD by active interventions. One approach aims to diminish transmission by instituting broad, cross-cutting water/sanitation/hygiene interventions to reduce the risk factors that result in fecally contaminated hands, food, and water and in allowing house flies to serve as mechanical vectors to carry enteric pathogens that can cause illness with small inocula (eg, *Shigella*). Examples of these interventions include household-based methods of treating water, refrigeration (to prevent pathogens in food and drink from growing to become potentially large inocula), washing hands with soap at critical points during the day (following defecation, prior to handling food, and before and after holding infants). Each of these interventions is estimated to diminish the incidence of diarrhea illness by 12%–25% [24–27], and each is presumed to be cross-cutting (ie, to diminish all enteric pathogens transmitted by a particular mechanism against which the intervention is directed).

The alternative strategy whereby the burden of MSD may be diminished, even without water, sanitation, and hygiene interventions that diminish the overall fecal burden in the environment, is to modify the immunologic status of the host from susceptible to immune by means of vaccination against specific pathogens. To pursue this strategy, one must first know the major agents responsible for MSD and their relative contribution to the overall MSD burden to prioritize what existing vaccines need to be implemented and what others need investments to be developed. It is in this context that the concept of AF is so potentially useful and important. While this might be straightforward in a study of diarrheal disease in an industrialized country setting, deciphering the data in a developing country project such as GEMS is daunting because approximately 85% of MSD cases can yield 1 or more enteropathogens, as can >70% of healthy controls, and a substantial proportion can yield multiple pathogens. Use of adjusted AF takes into account not only the prevalence of a pathogen of interest in controls as well as in cases, but also the presence of other pathogens besides the pathogen of interest, in both cases and controls.

Surprisingly, despite the fact that a large number of case/control studies have been carried out in developing countries to look for the predominant pathogens of MSD, the statistical analyses have only rarely utilized the concept of AF. In part this may have been due to lack of clear understanding of methods for adjustment for the presence of multiple pathogens [28]. However, the AF methodology, including the calculation of adjusted AFs, has undergone considerable development in recent decades. Several recent reviews have addressed subtleties in both the mathematical models and assumptions that underlie the use of AF [16, 28, 29]. One fundamental point to consider is that AF for an enteropathogen as a cause of MSD can be calculated either based on the distribution of exposure to the pathogen in the population [17] or the distribution of exposure in the cases [18]. In the GEMS analyses we employ the latter approach. We believe that calculation of AF, with adjustment for the presence of multiple pathogens among cases and controls, provides an appropriate approach for identifying the relative burden of diarrheal disease that could be eliminated through interventions against specific pathogens. We propose that this be adopted as a standard methodology (among others) for studies similar to GEMS, so that it will be possible to compare results of studies across time and geography, if other relevant case definition, selection, and laboratory methods are similar.

Notes

Acknowledgments. We are indebted to the following individuals who participated in 1 or more meetings of the GEMS Steering Committee on Biostatistical Issues and contributed their experience, wisdom, creativity, and advice: Lawrence Moulton, Janet Wittes, William Pan, and Peter Smith. We also thank Robert Black for valuable suggestions concerning GEMS statistical methodology.

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet **2010**; 375:1969–87.
2. Black RE, Lopez de Romana G, Brown KH, Bravo N, Bazalar OG, Kanashiro HC. Incidence and etiology of infantile diarrhea and major

routes of transmission in Huascar, Peru. Am J Epidemiol **1989**; 129:785–99.
3. Albert MJ, Faruque AS, Faruque SM, Sack RB, Mahalanabis D. Case control study of enteropathogens associated with childhood diarrhea in Dhaka, Bangladesh. J Clin Microbiol **1999**; 37:3458–64.
4. Tin A, Mar MN, Kyi KK, et al. Epidemiology and aetiology of acute childhood diarrhoea in Burma: a rural community survey. Trans R Soc Trop Med Hyg **1989**; 83:827–30.
5. De MP, Brasseur D, Hemelhof W, Kalala T, Butzler JP, Vis HL. Enteropathogenic agents in children with diarrhoea in rural Zaire. Lancet **1983**; 1:516–8.
6. Gascon J, Vargas M, Schellenberg D, et al. Diarrhea in children under 5 years of age from Ifakara, Tanzania: a case-control study. J Clin Microbiol **2000**; 38:4459–62.
7. Echeverria P, Taylor DN, Lexsomboon U, et al. Case-control study of endemic diarrheal disease in Thai children. J Infect Dis **1989**; 159:543–8.
8. Valentiner-Branth P, Steinsland H, Fischer TK, et al. Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. J Clin Microbiol **2003**; 41:4238–45.
9. Huilan S, Zhen LG, Mathan MM, et al. Etiology of acute diarrhoea among children in developing countries: a multicentre study in five countries. Bull World Health Organ **1991**; 69:549–55.
10. Levine MM. Enteric infections and the vaccines to counter them: future directions. Vaccine **2006**; 24:3865–73.
11. Levine MM, Kotloff KL, Nataro JP, Muhsen K. Impetus and rationale for the Global Enteric Multicenter Study (GEMS). Clin Infect Dis **2012**; 55(Suppl 4):S215–24.
12. Kotloff KL, Blackwelder WC, Nasrin D, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis **2012**; 55(Suppl 4):S232–45.
13. Korn EL, Graubard BI. Analysis of health surveys. New York: Wiley, **1999**:124–5.
14. Breslow NE, Day NE. Statistical methods in cancer research. Vol. I—The analysis of case-control studies. IARC Sci Publ **1980**; 202–5, 248–79.
15. Firth D. Bias reduction of maximum likelihood estimates. Biometrika **1993**; 80:27–38.
16. Hanley JA. A heuristic approach to the formulas for population attributable fraction. J Epidemiol Community Health **2001**; 55:508–14.
17. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum **1953**; 9:531–41.
18. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol **1974**; 99:325–32.
19. Sobel J, Gomes TA, Ramos RT, et al. Pathogen-specific risk factors and protective factors for acute diarrheal illness in children aged 12–59 months in Sao Paulo, Brazil. Clin Infect Dis **2004**; 38:1545–51.
20. Rothman KJ, Greenland S. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, **2008**.
21. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol **1985**; 122:904–14.
22. Rider PR. Variance of the median of small samples from several populations. J Am Stat Assoc **1960**; 55:148–50.
23. Sommerfelt H, Steinsland H, van der Merwe L, et al. Case/control studies with follow-up—constructing the source population to estimate effects of risk factors on development, disease and survival. Clin Infect Dis **2012**; 55(Suppl 4):S262–70.
24. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. BMJ **2007**; 334:782.
25. Clasen TF, Bostoen K, Schmidt WP, et al. Interventions to improve disposal of human excreta for preventing diarrhoea. Cochrane Database Syst Rev **2010**; CD007180.
26. Curtis V, Cairncross S. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. Lancet Infect Dis **2003**; 3:275–81.
27. Curtis V, Schmidt W, Luby S, Florez R, Toure O, Biran A. Hygiene: new hopes, new horizons. Lancet Infect Dis **2011**; 11:312–21.
28. Benichou J. A review of adjusted estimators of attributable risk. Stat Methods Med Res **2001**; 10:195–216.
29. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health **1998**; 88:15–9.

SUPPLEMENT ARTICLE

Data Management and Other Logistical Challenges for the GEMS: The Data Coordinating Center Perspective

Kousick Biswas,[1] Christina Carty,[1] Rebecca Horney,[1] Dilruba Nasrin,[2] Tamer H. Farag,[2] Karen L. Kotloff,[2] and Myron M. Levine[2]

[1]Cooperative Studies Program Coordinating Center, Department of Veterans Affairs, Perry Point, and [2]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

The Cooperative Studies Program Coordinating Center provided the data management, administrative, and statistical support to the Global Enteric Multicenter Study (GEMS). The GEMS study, the largest epidemiological study in the diarrheal disease area among children <5 years of age, was carried out in 4 African countries and 3 Asian countries. Given the geographical and geopolitical differences among the countries, the administration of a centralized data management operation was a major challenge. The sheer volume of the data that were collected, regular transfer of the data to a centralized database, and the cleaning of the same also posed some challenges. This paper outlines the details of the support that the data coordinating center provided and the challenges faced during the course of the study.

The Cooperative Studies Program Coordinating Center at Perry Point, Maryland, is one of 5 coordinating centers under Clinical Sciences Research and Development in the Department of Veterans Affairs, and specializes in providing data management, statistical, and administrative support to VA clinicians in the planning, conduct, and close-out of multisite clinical trials and epidemiological studies. In early 2006, the Center for Vaccine Development (CVD) of the University of Maryland School of Medicine approached the Perry Point Data Coordinating Center (DCC) for data management and other related services for the Global Enteric Multicenter Study (GEMS), an international epidemiological study of diarrhea in children <5 years of age, to utilize the Perry Point DCC's years-long experience in handling large-scale clinical studies.

With support from the Bill & Melinda Gates Foundation, GEMS was carried out in 7 countries: 4 sites in Africa (Basse, The Gambia; Kisumu, Kenya; Bamako, Mali; Manhica, Mozambique) and 3 sites in Asia (Kolkata, India; Mirzapur, Bangladesh; Karachi, Pakistan). GEMS began with a Health Utilization and Attitude Survey (HUAS) in each country, where approximately 1100 households were randomly sampled from either an existing or a newly initiated demographic surveillance system. After the completion of the HUAS, a 3-year case/control study was initiated in each country. In the case/control study, 660 children with moderate-to-severe diarrhea (cases), along with 660 matched children without diarrhea (controls), were recruited in each of 3 age strata (0–11, 12–23, and 24–59 months) in each country. During the 3-year case/control study, a shorter version of the original HUAS ("HUAS-Lite") was performed 2–3 times per year where 1100 households were randomly sampled from the respective demographic surveillance system for each round completed. The study used a total of 20 case report forms (CRFs). These included

Correspondence: Kousick Biswas, PhD, Cooperative Studies Program Coordinating Center, Department of Veterans Affairs, PO Box 1010, Perry Point, MD 21902 (kousick.biswas@va.gov).

Clinical Infectious Diseases 2012;55(S4):S254–61

DOI: 10.1093/cid/cis755

1 CRF for the original HUAS, 1 CRF for the HUAS-Lite, 9 clinical/epidemiological CRFs (5 for cases, 4 for controls), 6 laboratory CRFs, and 3 verbal autopsy CRFs. The interviews for HUAS and for HUAS-Lite rounds and the data collection for the case/control study were conducted by locally employed and trained health workers in each country. To satisfy the need of the health workers who spoke native languages, the CRFs and the informed consent forms were translated and printed in different languages. For example, Mali used CRFs in French, Mozambique in Portuguese, and Kenya in dual languages (Dholuo and English). Bangladesh used informed consent forms in Bengali and Pakistan in Urdu.

PERRY POINT DCC SUPPORT

Once contracted by the CVD and approved for participation by the VA central office, the DCC established a core team for the GEMS. The core team was comprised of a team lead (also a biostatistician), a statistical programmer, a project manager, 2 data managers, and 4 computer assistants. The team lead and the project manager provided the administrative support, the biostatistician and the statistical programmer provided the statistical support, and the data managers and the computer assistants provided the data management support for the study.

The Perry Point DCC provided the following services for the GEMS:

Data Management Support

(a) Implementation of a data flow model where data collection would take place at the individual countries and data would be sent to the DCC periodically where the study master database(s) would be established, maintained, and managed.

(b) Selection of a standardized data management platform that would work seamlessly in 8 different countries in 3 continents (7 participating countries in 2 continents and the United States as the central hub for data management) with a significant variability in technical support and other related logistical support.

(c) Design standardized CRFs that adhere to specifications required by the selected data collection software.

(d) Generate paper CRFs in 4 languages (English, French, Portuguese, and Dholuo).

(e) Generate paper informed consent forms in 3 languages (English, Bengali, and Urdu).

(f) Establish a standardized data transfer protocol between the participating countries and the DCC.

(g) Design a data quality control (QC) protocol where "data QC" reports would be generated and sent to the participating countries periodically.

(h) Design a data accountability protocol where "missing CRF" reports would be generated and sent to the sites periodically based on an "expected CRF" algorithm.

(i) Establish a study-specific numbering protocol for the HUAS households, case/control children, and their laboratory samples.

(j) Establish a data management handbook with instructions for CRF completion, addressing the QC and missing form reports, etc.

(k) Build and maintain close contacts between the data management workgroups in the participating countries and the data management group at the DCC using conference calls, emails, etc.

Administrative Support

(a) Implementation and tracking of "CRF request forms" from each participating country.

(b) Printing and shipping of CRFs for each country for the entire duration of the study.

(c) Tracking of all regulatory documentation from each participating countries, including institutional review board/ethics board approvals, Federal Wide Assurance numbers, translation certificates, etc.

(d) Setup of regular meetings between the CVD core group and the DCC.

(e) Maintain required minimum staff to ensure execution of the DCC support.

(f) Participate in annual or other meetings during the study.

Statistical Support

(a) Generate weekly tables on eligibility and enrollment.

(b) Generate monthly aggregate tables on variables (either original or constructed) as requested by the CVD core group.

(c) Generate analytic datasets for analysis purposes as requested by the CVD core group.

(d) Perform statistical analysis based on an established statistical analysis plan.

(e) Participate in statistical workgroup meetings.

DATA COLLECTION TOOL DATAFAX: AN OVERVIEW

DataFax was chosen as the data collection tool for GEMS by the data management group from DCC. The primary objective of DataFax is to automate the collection and processing of paper case report forms, and ultimately improving the timeliness and quality of the study database. The specific design objectives are as follows:

- **Use simple technology in the clinical sites.** Clinical sites (or participating countries) can send CRFs to the DCC in TIFF

or PDF format using either an ordinary fax machine (from any standard G3 fax machine), an Internet fax machine, email, or secure file transfer protocol (SFTP). Data can be added via raw data entry into data screens (from paper CRFs), or by importing ASCII data files (eg, from a central laboratory).

- **Computerize the receipt, logging, and filing of CRFs.** When data are submitted to DataFax, the software reads information embedded in barcodes on the individual CRF pages and routes them to the appropriate study database. Each CRF image is assigned a fax ID and is placed in a queue for validation. Once validation is complete, all images and data records are stored electronically in a secure location on optical or magnetic disc.
- **Automatically generate an initial data record.** DataFax reads data boxes (Xs, handwritten numbers, and visual analog scales) to create an initial data record as the starting point for the clinical review and data validation process.
- **Provide split screen review of CRFs and the corresponding data records.** All CRFs and initial data records are reviewed on screen to complete data entry, make corrections, and flag problems (eg, missing data).
- **Automate the QC process.** Problems detected on the CRFs received by the DCC are flagged using QC notes (electronic sticky notes), which are automatically formatted into standard QC reports for transmission by fax or email to the clinical sites.
- **Automate work flow management.** CRFs are stamped with a validation level at each CRF review and data processing stage.

DataFax does not work with arbitrary CRFs. Study CRFs have to be designed by the DCC to adhere to DataFax specifications, which include the following:

- Bar coding must be placed at the top of each CRF page to identify the study, the CRF plate (page), and optionally the sequence (or visit) number.
- All pages must be US letter or A4 size and oriented vertically (portrait) not horizontally (landscape). The DCC selected standard US letter size for the GEMS study.
- All boxes designed for numerical data and spacing must conform to DataFax standards.

Faxed/scanned/emailed CRFs are automatically indexed upon receipt (by study number, CRF page number, and optionally by visit number) from barcodes printed at the top of each CRF page. The remaining fields on each CRF page are processed by the intelligent character recognition (ICR) software, which reads numeric, date, check, choice, and visual analog scale fields to create an initial data record ready for subsequent validation by the data management staff. Text fields are not read by the ICR software and must be entered manually when the record is validated.

Each newly received CRF page and the corresponding data record created by the ICR software are reviewed by data management staff using the DataFax validation tool. Data management staff flag any CRF problems (eg, missing data), using pop-up QC notes, during validation. Lookup tables provide standardization of queries to be sent to clinical investigators. Preprogrammed edit checks will detect inconsistencies within forms, across forms, across visits, and even across study participants, if necessary.

A QC report program formats all QC notes for each clinical site into a clear, compact report identifying all outstanding CRF problems and clarification requests. Missing pages and overdue visits may be included in QC reports. Each QC report may also include a scheduling summary for all participants at the clinical site including, entry date, date of last visit, and target date for the next scheduled visit. The QC reports may be faxed and/or emailed to clinical sites at scheduled times.

Corrected CRFs, re-sent from the clinical sites, are automatically identified on arrival for revalidation, entry of corrected fields, and resolution of QC notes using the DataFax validation tool. All versions of each CRF page (and all versions of the corresponding data records) are retained for subsequent review, but only 1 version of each CRF page is flagged as the primary (good) copy and linked to the primary data record.

Full journaling identifies all changes made to the database by user, date, and time. A QC database tracks all data clarification queries by problem type (eg, missing data, illegal value) and current status (resolved or outstanding). Audit trail reports show all changes made to the database, at data record and individual data field levels by user, date, and time, including history of QC notes.

DATA TRANSFER PROTOCOL

The data flow model that was adopted for GEMS was as follows:

1. Data were collected in the field in each participating country using paper CRFs in the appropriate language with appropriate barcodes.
2. Completed CRFs were scanned and saved as TIFF or PDF files using a scanner with prespecified resolution settings (each country purchased a scanner locally based on the supplied specifications by the DCC) to ensure readability by the DataFax software.
3. The TIFF and PDF files were electronically transferred to the DCC at an agreed-upon interval.

(a) Three different transfer platforms were used for GEMS:

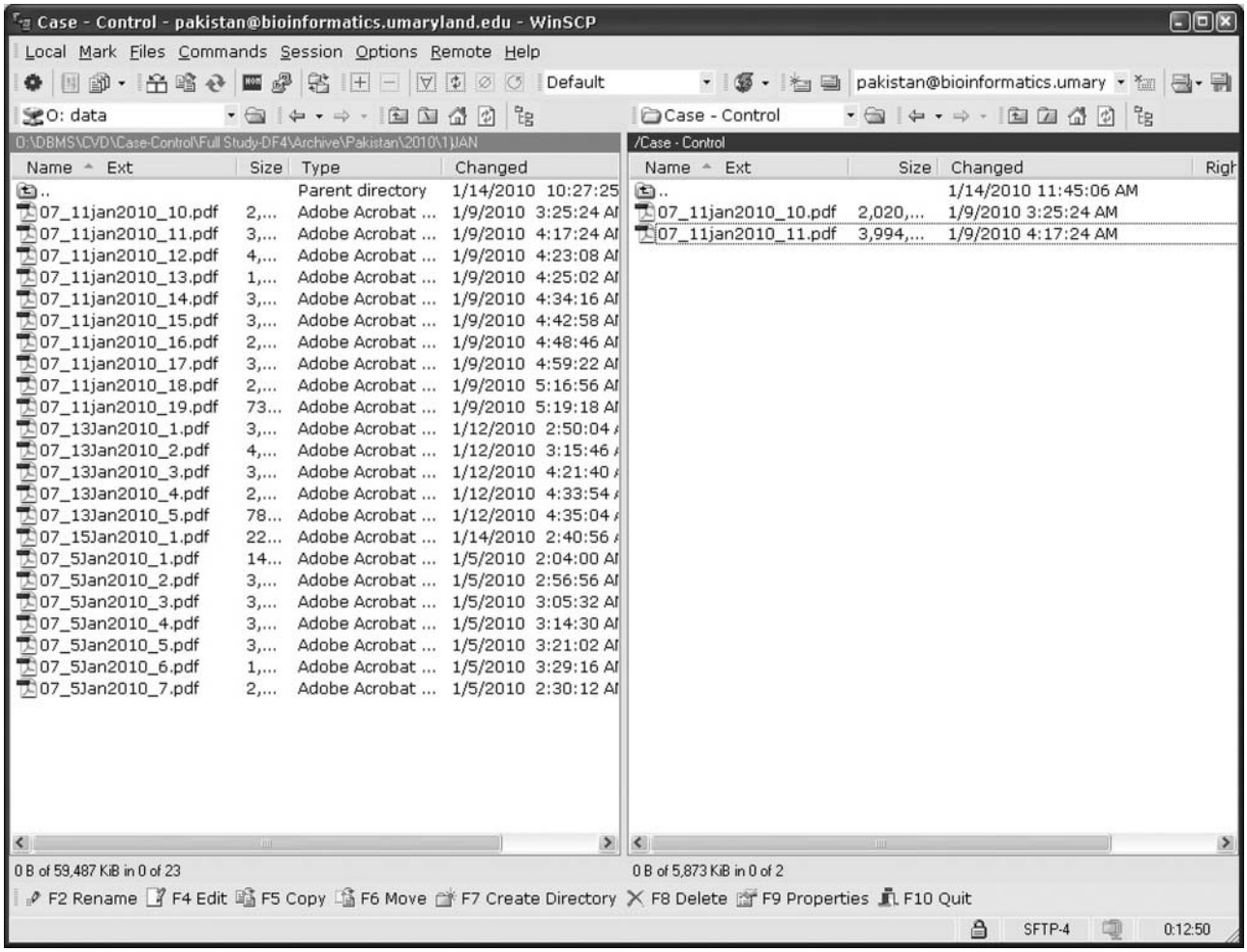

Figure 1. Structure of a typical FTP (file transfer protocol) account.

(i) Email: In the beginning of the study, the participating countries were requested to send the scanned files as attachments via emails, but the size limitations for the attached documents as outlined by the VA exchange email server (5 MB) made this mode very time consuming and difficult to track as it was requiring the sites to send multiple emails to be within the acceptable file size limit.

(ii) Microsoft Groove: Collaborative workspaces were created for each country where the sites could post their scanned files for DCC to retrieve from. Each country's folder structure in their respective workspaces was set up by the DCC staff to ensure ease of posting and retrieval of files. DCC had access to all of these workspaces, but each country's access was restricted to its own workspace only. DCC staff also regulated/controlled access to these workspaces by the site staff. DCC staff deleted the posted files regularly once the files were retrieved from the workspaces to keep the workspace synchronization times under control. These workspaces were working very well for data transfer until a decision was made by the GEMS executive committee to use these workspaces as archives for the submitted CRFs. At the end of the second year of the study the size of each workspace exceeded the size limit of Groove (2 GB), which made the workspaces unusable.

(iii) SFTP server: As an alternative solution, accounts were created on an SFTP server accessible by each participating country for data transfer. The screenshot in Figure 1 shows an account for one of the participating countries.

4. The scanned files, when received at the DCC, were routed to the respective form queues based on the barcodes placed on top of each page of each CRF.

(a) For the ease of data management for GEMS, 4 separate databases (and thus 4 form queues) were maintained—1 for HUAS (included data from original HUAS and the HUAS-Lite rounds), 1 for case registration (CRF 2), 1 for the rest of the case/control study (included all the clinical/epidemiological and the laboratory CRFs), and 1 for the verbal autopsy data.

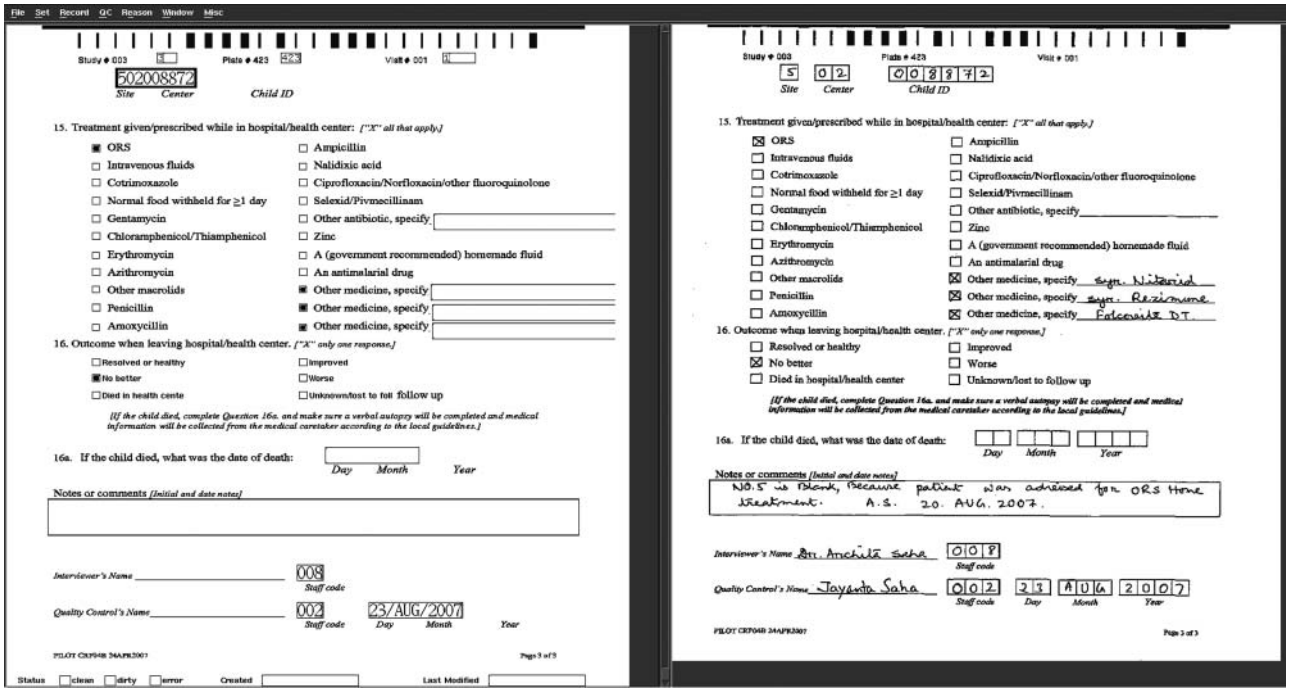

Figure 2. Split-screen validation in DataFax.

5. Before submission to the respective master databases, the data management group at the DCC validated the forms on a split screen. The right half of the screen displays the actual scanned image of a CRF page and the left half of the screen displays the software's ICR interpretation of the same page (Figure 2).

6. Based on the discrepancies/errors/illegible entries identified during split-screen validation, QC notes are added and reports were generated and sent back to the countries for corrections/explanations.

7. The data management workgroup at each country rescanned the corrected pages of the CRFs and posted them in an appropriate folder on the SFTP server for the DCC to retrieve. Once received at the DCC, the CRF pages were revalidated before being submitted to the respective databases.

8. Once the data were cleaned, the datasets were shared with each country and the CVD core group intermittently.

9. After the final data lock, entire datasets from the respective countries were sent for final analysis.

NUMBERING PROTOCOL

For GEMS, a unique identification number was assigned to each household surveyed for the HUAS (original and the Lite rounds), to each child enrolled as a case or control for the case/control study, and to each laboratory sample that was collected from each child enrolled. To achieve this task, a numbering protocol was established by the DCC and the details are given below:

- **HUAS**
  - ID: 5 numbers divided into 2 sections
    - Section 1: Position 1 = Country # (1–7, 9 possible)
    - Section 2: Positions 2 to 5 = Survey # (0001–9999)
    - Entire ID numbers are preprinted on CRFs
    - Child ID numbers assigned sequentially within each country
      - Ex. 10001 = Country 1, Sequential # 0001
- **HUAS Lite**
  - ID: 7 numbers divided into 2 sections
    - Section 1: Position 1 = Country # (1–7, 9 possible)
    - Section 2: Positions 2 to 7 = Survey # (000001–999999)
    - Entire ID numbers are preprinted on CRFs
    - Child ID numbers assigned sequentially within each country
      - Ex. 1000001 = Country 1, Sequential # 000001
- **CASES AND CONTROLS**
  - ID: 9 numbers divided into 3 sections
    - Section 1: Position 1 = Site # (1–7, 9 possible)
    - Section 2: Positions 2 to 3 = Center # (99 possible for each country)

        - Section 3: Positions 4 to 9 = Sequential # (000001–999999)
    - Case ID number transcribed from the Registration Log (CRF 2)
        - Sequential numbering from 000001 to 899999
        - Ex. 101000001 = Country 1, Center 01, Sequential # 000001
        - Ex. 102000025 = Country 1, Center 02, Sequential # 000025
    - Control Patient ID
        - Sequential numbering from 900001 to 999999
        - Ex. 101900001 = Country 1, Center 01, Sequential # 900001
        - Ex. 202900050 = Country 2, Center 02, Sequential # 900050
- **SPECIMEN IDs**
    - ID: 6 numbers divided into 2 sections
        - Section 1: Position 1 = Country # (1–7)
        - Section 2: Positions 2 to 6 = Sequential numbering from 00001 to 99999
        - Ex. 100001 = Country 1, Sequential # 00001

DCC CHALLENGES

The following sections outline the challenges that DCC experienced and the lessons learned at different phases of GEMS.

One of the major challenges that the data management group experienced was to identify and implement a data management platform that would work seamlessly in 8 different countries (7 participating countries and the US as the centralized data hub). DataFax served very efficiently as a centrally managed data management system supporting standardized CRFs with a standardized data transfer protocol for the GEMS. DCC supplied the technical specifications of scanners, which were purchased/installed/managed locally at each country and used to scan and generate TIFF or PDF files with a specified resolution setting for validation purposes at the DCC. The other major challenge the data management team at DCC faced was the use of different language CRFs in some countries. India, Bangladesh, Pakistan, and The Gambia used CRFs in English but Mali, Kenya, and Mozambique used CRFs in the languages spoken in those countries. The challenge was to maintain the English CRF format (length, number of questions on each page, location of boxes for responses on each question) intact on each page of a given CRF while the translations in respective languages (French for Mali, Portuguese for Mozambique, and Dholuo for Kenya) were overlaid either in place of corresponding English questions or in combination. The validation screens at the DCC were maintained in English for standardization purposes.

The validation process also posed some challenges to the DCC staff. Because DataFax uses an ICR technology to recognize "X" and numerals, it did not recognize handwritten notes. The computer assistants at the DCC needed to manually type the handwritten responses. Even though the number of handwritten responses was limited, illegible handwriting (often in languages other than English) posed challenges.

Other challenges, related to the data management processes that the DCC staff encountered, are outlined below:

(a) Printing and shipping of CRFs to ensure smooth conduct of data collection in each participating country: Because a paper-based data management system was chosen for GEMS, an enormous amount of CRFs needed to be printed and shipped to the participating countries. Since DataFax was very restrictive about the paper size, the decision was made to print all the CRFs in US standard letter size. Country-specific mailing times and customs issues posed additional challenges toward maintaining the timeline. Moreover, receipt of CRF requests from the countries in a timely manner was also a crucial element in maintaining the timeline.

(b) CRF submission issues:

(i) Not using the correct/recommended settings for the scanner by the participating countries during scanning of the paper CRFs (this posed challenges in reading the CRFs correctly by the DataFax software);

(ii) Not including the file submission tracking log with each submission of CRFs;

(iii) Not following the suggested naming conventions of the submitted files.

(c) Data cleaning challenges:

(i) Overlooking some of the identified errors on a QC report (in these situations the QC reports with the same errors were sent back repeatedly for correction);

(ii) Not following the "Resubmission of CRFs with Corrections" protocol while sending back the corrected CRFs to the DCC;

(iii) Long-overdue CRFs.

To overcome some of these challenges, the DCC data management staff scheduled numerous conference calls with each participating country to enforce various standardized protocols that DCC established for GEMS and also to help the site personnel to understand the contents of the QC reports.

SOME DATA ON THE SCALE OF GEMS

GEMS is the largest international, epidemiological study executed in the diarrheal disease area in Asia and in sub-Saharan Africa. GEMS amassed a huge volume of data. In the original round of HUAS, a total of approximately 7000 households

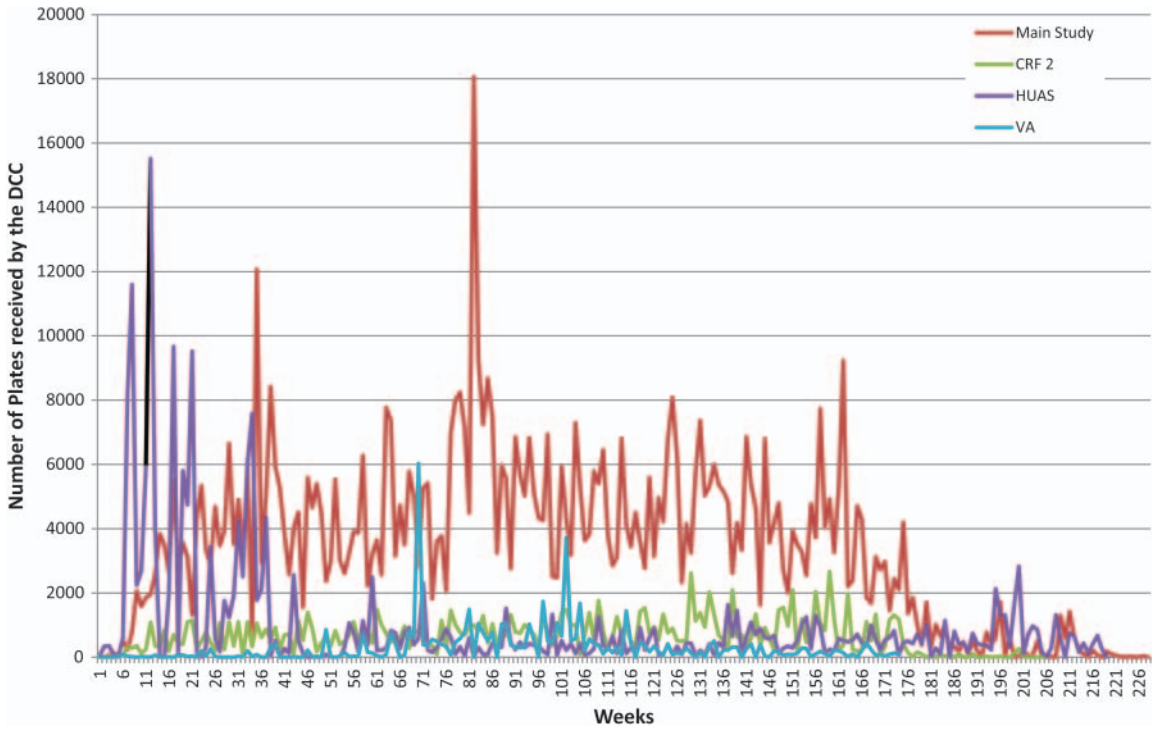

Figure 3. The number of case report forms (CRFs) received in the 4 databases (main study, CRF2, HUAS, and Verbal Autopsy) of Global Enteric Multicenter Study. Abbreviations: CRF, case report forms; HUAS, Health Utilization and Attitude Survey; VA, Verbal Autopsy; DCC, Perry Point Data Coordinating Center.

were surveyed. On average, each country completed 6 rounds of HUAS-Lite surveys during the 3-year case/control study, with the exception of 2 countries. In total, 30 000–35 000 households were surveyed during the HUAS-Lite rounds. For the case/control study, approximately 27 000 children were enrolled in the 7 participating countries. In total, the DCC validation team has processed about 1.5 million pages of CRFs for GEMS. Figure 3 illustrates the volume of CRFs received over the first 130 weeks of the study and Figure 4 provides a perspective of GEMS' scale compared to other studies conducted at DCC.

DISCUSSION

The GEMS project was an enormous undertaking from the DCC perspective, not only because of its scale but also for its involvement with multiple countries from 3 continents with very diverse cultural, social, and technological backgrounds. It was a valuable experience for the DCC staff, who learned from dealing with a group of talented, diligent, and focused individuals from the 7 countries. DCC staff was successful in building up close and mutually respectful working relationships with the data management staff from all the participating countries. The use of standardized case report forms for data collection, utilization of standardized data management software, and the application of numerous standardized data management procedures all helped the data management staff at the DCC to successfully collect and clean a huge amount of data in a timely fashion.

Arguably, perhaps the most onerous challenge for DCC was to print and ship enormous amounts of paper CRFs on a tight schedule to ensure continuity in data collection at the sites. An electronic data management/transfer system with electronic CRFs could have eliminated this huge and expensive logistical challenge. On the other hand, at least initially, the variability in availability of high-speed Internet technology, the varying degrees of expertise of local staff in some countries to manage a sophisticated system, and the challenge of implementing and managing an electronic system remotely were some of the reasons for DCC to go with a more conventional system.

As the data coordinating center for large cooperative clinical trials sponsored by the Department of Veterans Affairs, over the years our unit has gained wide recognition for the large and complex datasets that we have managed. Yet GEMS is unique in regards to the number of plates that the DCC staff processed over the duration of the study, surpassing all previous experience. Indeed, in comparison with other large studies completed at the DCC, GEMS was almost 10 times larger in

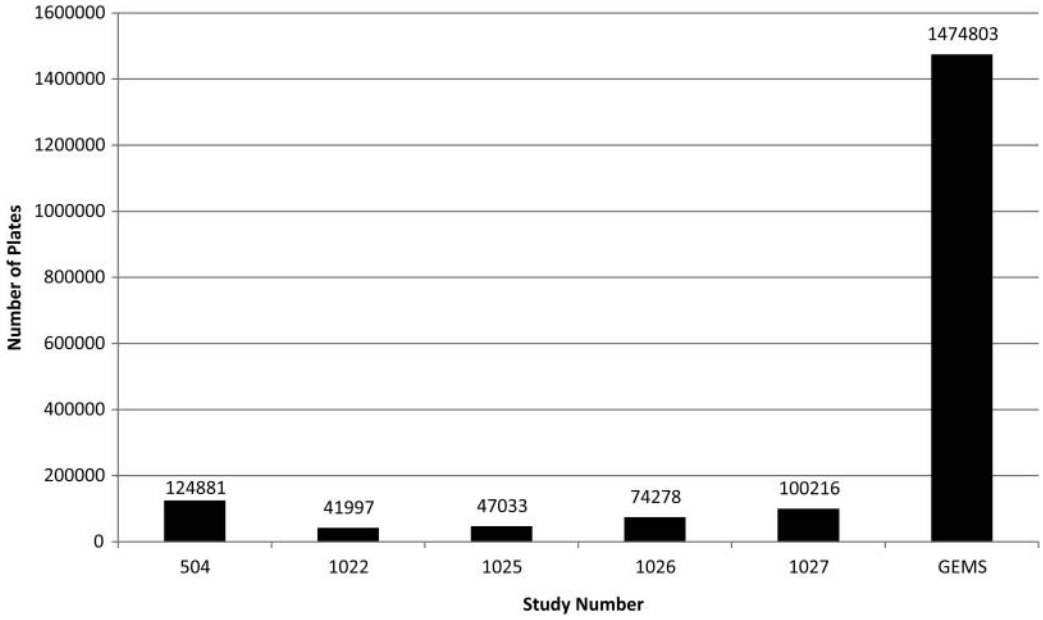

Figure 4. The number of plates received for Global Enteric Multicenter Study versus other studies completed by the Data Coordinating Center.

regards to the volume of data. We believe that the insights and experiences that we have described in this paper should be helpful to other research consortia undertaking projects that generate enormous datasets and that must transfer those data expeditiously from field sites (including some very rural sites) in multiple developing countries to a central data coordinating center.

Notes

Acknowledgments. The enormous task of implementing and conducting the study from the Perry Point Data Coordinating Center (DCC) perspective would not have been possible without the talented and hardworking DCC staff. The DCC team was comprised of the following individuals in addition to the first author as the team lead and the biostatistician: project managers—Barbara Yndo, Karen Hessler, Kristy Tomlin; data managers—Steven Berkey, Christina Carty; statistical programmer—Rebecca Horney; computer assistants—Ellen Sterrett, Christopher Crayton, Veronica Debity, Julie Lowe, Carla Davis, Tangra Cole.

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

SUPPLEMENT ARTICLE

Case/Control Studies With Follow-up: Constructing the Source Population to Estimate Effects of Risk Factors on Development, Disease, and Survival

Halvor Sommerfelt,[1,2] Hans Steinsland,[1,3] Lize van der Merwe,[4,5] William C. Blackwelder,[6] Dilruba Nasrin,[6] Tamer H. Farag,[6] Karen L. Kotloff,[6] Myron M. Levine,[6] and Håkon K. Gjessing[7,8]

[1]Centre for International Health, University of Bergen, [2]Division of Infectious Disease Control, Norwegian Institute of Public Health, Oslo, [3]Department of Biomedicine, University of Bergen, Norway; [4]Biostatistics Unit, Medical Research Council, and [5]Department of Statistics, University of Western Cape, Cape Town, South Africa; [6]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; [7]Division of Epidemiology, Norwegian Institute of Public Health, Oslo, and [8]Department of Public Health and Primary Health Care, University of Bergen, Norway

If individuals in a case/control study are subsequently observed as a cohort of cases and a cohort of controls, weighted regression analyses can be used to estimate the association between the exposures initially recorded and events occurring during the follow-up of the 2 cohorts. Such analyses can be conceptualized as being undertaken on a reconstructed source population from which cases and controls stem. To simulate this population, the cohort of cases is added to the cohort of controls expanded with the reciprocal of the case disease incidence odds (the sampling weight) to include all individuals in the source population who did not develop the case disease. We use a simulated dataset to illustrate how weighted generalized linear model regression can be used to estimate the association between an exposure captured during the case/control study component and an outcome that occurs during follow-up.

By including a larger fraction of individuals in a source population who develop a disease than of those who do not, case/control (CC) studies are more efficient than the corresponding cohort studies in obtaining measures of association between exposures and disease risk [1–3]. With decreasing disease incidence, this sampling fraction decreases, and the relative efficiency of CC studies increases. In CC studies nested inside a defined cohort, the sampling fraction can be calculated directly as the number of controls, that is, the disease-free individuals in whom exposures are recorded, divided by the total number of individuals who did not develop the disease during the course of the cohort study [2–5].

Even CC studies that are not undertaken within a defined cohort can be conceptualized as being nested in a source population [2]. This population, or the underlying, "hypothetical" cohort, is elusive because it is neither captured in a roster nor followed to record outcomes. In CC studies that reuse data for measuring associations between exposures and an outcome other than that defined by being a case, the occurrence of the case disease can be used to calculate weights for appropriate regression analyses [6]. Once the disease risk is estimated, subsequent follow-up of CC study participants in a cohort of cases (CoCa) and a cohort of controls (CoCo) enables us to measure the association between exposures recorded at recruitment into the study and an outcome during follow-up, such as growth, disease, or death. Provided there is an association between an exposure and becoming a case, and

Correspondence: Halvor Sommerfelt, MD, PhD, Centre for International Health, University of Bergen, PO Box 7804, 5020 Bergen, Norway (halvor.sommerfelt@cih.uib.no).

Clinical Infectious Diseases 2012;55(S4):S262–70

DOI: 10.1093/cid/cis802

cases have a higher risk of outcomes measured during follow-up than controls, CC studies with follow-up (CCF) are more efficient in identifying an association between the exposure and such outcomes than the corresponding cohort study, because the exposure is condensed in the CoCa.

We created an imagined population with a known exposure, case disease occurrence, and outcome distribution in order to present a conceptual framework of CCF data analysis using what we call the reconstructed population method (RPM). We then show how this framework can be translated into weighted regression analysis.

DECOMPOSING THE POPULATION AND THEN RECONSTRUCTING THE UNDERLYING COHORT FROM THE CASES AND THE CONTROLS

Let us imagine a population with N individuals where we recruit 1 control per case into a CC study and follow up the CoCa and the CoCo. The exposure (E) and outcome (O) are distributed as shown in Figure 1.

Our incident cases are recruited into this CCF study within a short time window after the onset of a case-defining event (D); those who do not develop D within that time window are noncases (NC). The population was generated using functions (given in the Supplementary Appendix) that describe how E influences D, and how E and D separately and in combination, influence O (Figure 2).

Because our study recruits an equal number of cases and controls, n, the number of NC in the population is N-n. Because n of all NC are recruited into the CoCo, the sampling fraction is calculated as n/(N-n), and the sampling weight as (N-n)/n [7]. In our example, where n = 2400 and N = 100 000, the sampling fraction is 2400/(100 000–2400) = 0.02459, the corresponding sampling weight (100 000–2400)/2400 = 40.67.

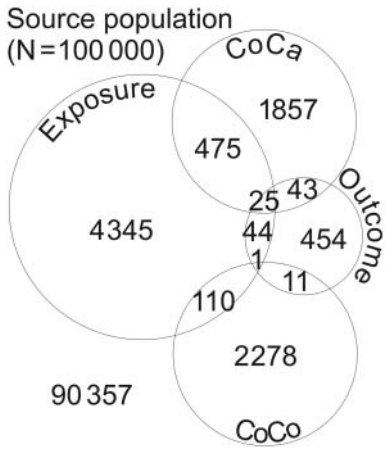

Figure 1. Venn diagram showing the distribution of 2400 cases and 2400 controls in relation to an exposure and an outcome in a population of 100 000 individuals. The numbers were generated using functions found in the Supplementary Appendix. Abbreviations: CoCa, cohort of cases; CoCo, cohort of controls.

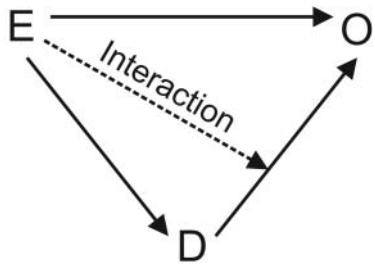

Figure 2. Schematic presentation of associations between an exposure (E), a case disease (D), and an outcome (O) in a population, where arrows indicate the direction of causality.

The relative risk (RR) of experiencing O given E for the whole population, in the CoCa, among the NC, and in the CoCo is shown in Table 1, rows A–D. The slight difference in RR between NC (Table 1, row C) and the CoCo (Table 1, row D) is an artifact of rounding.

The odds ratio (OR) describing the association between E and getting D (ie, becoming a case) can be calculated from Table 1, row E (derived from Table 1, rows B and D), which distributes E among the cases (D^+) and the controls (D^-).

We now make a shift to the real world of epidemiology where only the CCF study represented by the CoCa (Table 1, row B) and the CoCo (Table 1, row D) is known. It is only *conceptually* nested in the source population (Figure 1). When analyzing the CCF study with the RPM, the exposure-outcome distributions in the CoCa and the CoCo should be identical but because they represent samples of our population, estimated associations should be provided with confidence intervals (CIs) (Table 1, rows F and G).

As an estimate for the association between E and O in the source population, it may seem tempting to ignore the CC sampling scheme and simply calculate RR on the combined data of the 4800 individuals in CoCa and CoCo (Table 1, row H). This corresponds to what Jiang et al lists as the first ad hoc approach to secondary analysis of CC data [8]. However, this approach assumes that D is conditionally independent of O given E, ie, when none of the effect of E on O is mediated by D. When getting D, on the other hand, does change the risk of O, this approach yields an unbiased estimate of the association between E and O only when the ratio of cases to NC in the source population is 1:1, ie, when D risk is 50%. In many situations, including in the Global Enteric Multicenter Study (GEMS) [9, 10], not only may E increase the incidence risk of D, this risk is usually much lower than 50%, and such

Table 1. Two-by-Two Tables Showing Distributions of Exposure (E) and Outcome (O) or Disease Defining Case Status (D) as a Basis for the Conceptual Framework of the Reconstructed Population Method

			O					
			+	–	Total	Risk	RR	95% CI
A	Source Population							
	E	+	70	4930	5000	0.014	2.6	
		–	508	94 492	95 000	0.005		
B	CoCa							
	E	+	25	475	500	0.050	2.2	
		–	43	1857	1900	0.023		
C	NC							
	E	+	45	4455	4500	0.010	2.0	
		–	465	92 635	93 100	0.005		
D	CoCo							
	E	+	1	110	111	0.009	1.9	
		–	11	2278	2289	0.005		
			D					
			+	–	(Odds)	(OR)		
E	CC study							
	E	+	500	111	4.505	5.4		
		–	1900	2289	0.83			
			O					
			+	–				
F	CoCa							
	E	+	25	475	500	0.050	2.2	1.4–3.6
		–	43	1857	1900	0.023		
G	CoCo							
	E	+	1	110	111	0.009	1.9	.24–14.4
		–	11	2278	2289	0.005		
H	CoCa + CoCo							
	E	+	26	585	611	0.043	3.3	2.1–5.2
		–	54	4135	4189	0.013		

Abbreviations: CC, case/control; CI, confidence interval; CoCa, cohort of cases; CoCo, cohort of controls; D, case-defining illness; NC, noncases; OR, odds ratio; RR, relative risk.

an approach would accordingly overestimate the strength of the association between E and O (Table 1, row H).

Another approach, which is suggested by Nagelkerke et al, is to base the estimates only on the 2400 CoCo individuals (Table 1, row G) [4]. Jiang et al argues that this, in what they call the second ad hoc approach, may be approximately valid when D is rare [8], but emphasizes, just as do Reilly et al [6], that it is inefficient because it discards the case data. If there is an interaction between E and D on O, ie, when the association between E and O differs between CoCa and CoCo individuals, the bias may be substantial and even more unpredictable.

A third approach is to calculate RRs for the CoCa and for the CoCo, and, if there is no interaction between E and D on O, report the average of the 2 RRs using Mantel-Haenszel stratified analysis. This corresponds to Jiang et al's third ad hoc approach where the combined analysis of CoCa and CoCo individuals is adjusted for D [8]. This approach, which gives an RR estimate of 2.2 (95% CI, 1.4–3.5) in our example, not only disregards the fact that cases are oversampled (see Table 1, row H and the first ad hoc approach) but also de facto removes the effect of E on O that operates through, ie, is mediated by, D.

To use CCF data to estimate the association between E and O in a given population, we need to perform the analysis on the population reconstructed from the CoCa plus the NC. The sampling fraction needed to estimate NC cannot be calculated directly, but must be derived from an independent source of D incidence risk. Thus, if R is the incidence risk of D in the time window during which cases are recruited, and because we

Table 2. Reconstructing the Population

			O				
			+	–	Total	Risk	RR
A	rNC						
	E	+	1 × 40.67 = 40.67	110 × 40.67 = 4473.33	4514	0.013	1.9
		–	11 × 40.67 = 447.33	2278 × 40.67 = 92 639.67	93 086	0.005	
B	Reconstructed population[a]						
	E	+	40.67 + 25 = 65.67	4473.33 + 475 = 4948.33	5014	0.013	2.5
		–	447.33 + 43 = 490.33	92639.67 + 1857 = 94 496.67	94 986	0.005	

Two-by-two tables showing distributions of exposure, outcome, and disease that defines case status in the reconstructed noncases and in the reconstructed source population.

Abbreviations: D, case-defining illness; E, exposure; O, outcome; rNC, reconstructed noncases; RR, relative risk.

[a] rNC + cohort of cases.

assume equal numbers of cases and controls, the sampling fraction of controls is proportional to the corresponding incidence odds, ie, R/(1 – R). If we assign a weight of 1 to the cases, the sampling weight of the controls is its reciprocal, (1 – R)/R.

In our example, let us assume that the D incidence risk, derived from a perfectly representative survey in the population, is 0.024. To reconstruct the population's NC, we multiply the number of individuals in the CoCo with the reciprocal of its corresponding incidence odds, the sampling weight, ie, 40.67, to obtain the reconstructed number of exposed and unexposed noncases (rNC) (Table 2, row A). We can then estimate the association between E and O in our reconstructed population consisting of the CoCa plus the rNC (Table 2, row B).

The difference in cell numbers between the imagined (Table 1, row A) and this reconstructed population is an artifact of the rounding we undertook to generate the CoCo (Table 1, row D). We do not include a 95% CI for this RR estimate because the sampling error should be derived from the CoCa and CoCo, not from the reconstructed population.

To explain the difference between the RR in the combined CoCa and the CoCo (Table 1, row H) and that in the reconstructed population (Table 2, row B), and to provide a transition into regression analysis of such data, Figure 3 illustrates how the weighting of the data influences the estimated effect of E on O.

ANALYSIS OF CCF DATA USING WEIGHTED REGRESSION ANALYSIS

As a more versatile analytic approach than that depicted in the previous section, we will now describe a weighted generalized linear model (GLM), illustrated graphically in Figure 3*B*). It is based on a dataset containing individual records for the

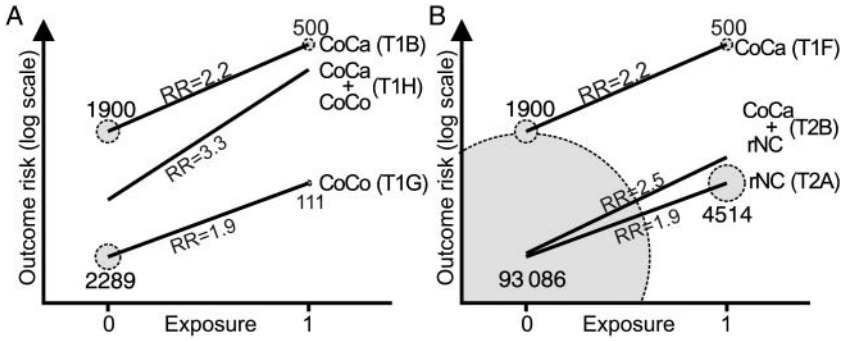

Figure 3. Regression lines reflecting the relative risk for an outcome during follow-up for (*A*) the cohort of cases (CoCa) + the cohort of controls (CoCo) and (*B*) the reconstructed population (CoCa + noncases that have been reconstructed from the CoCo × sampling weight [rNC]). The data underlying each line corresponds to the 2 × 2 tables in Table 1 and Table 2, so that (T1B) is the 2 × 2 in row B of Table 1, and (T2A) is the 2 × 2 table in row A of Table 2. Notice that the change in weights, or individuals, between (*A*) and (*B*) alters the end-point positions, and thus the slope of the middle line. (*A*) depicts the ill-advised approach to analyze the combined CoCo and CoCa data (Table 1, row H). The area of each circle is proportional to the number of exposed (Exposure = 1) and unexposed (Exposure = 0) individuals in the CoCa and the CoCo. (*B*) depicts the reconstructed population method (Table 2, row B). The area of each circle is proportional to the number of exposed and unexposed individuals in the CoCa and the rNC. Abbreviations: CoCa, cohort of cases; CoCo, cohort of controls; rNC, noncases that have been reconstructed from the CoCo × sampling weight; RR, relative risk.

4800 individuals in the CCF, with variables indicating E and O status, as well as the above-mentioned sampling weight. Our Supplementary Appendix contains instructions for the use of R-functions and a spreadsheet to generate the data we have used in this paper and data with other underlying associations between E, D, and O.

We further address how weighted GLM can be used to depict interactions between E and D on O, and show how to estimate the extent to which getting D mediates an effect that E has on O. We have chosen to use Stata version 12.1 (Stata Corp) to illustrate the analyses, but other statistical software, such as R (The R Foundation for Statistical Computing; www.r-project.org) can also be used for the weighted regression analysis, notably using survey weights [11, 12].

The 3 above-mentioned ad hoc approaches disregard the fact that in CC studies where D incidence risk is <50%, the cases are oversampled [4]. Several of the cited papers advise weighting the cases and the controls according to their relative probability of being sampled into the study [4, 6, 8]. For the analysis of CCF studies where 1 control is included per case, and the sampling weight for each individual in the CoCa, ie, for each case, is set to 1, the weight for the controls is then (1 – R)/R, as described above. If there are n cases and m controls, the weights are 1 for cases and (n/m)(1 – R)/R for controls. Specifying sampling weights (called *pweight* in Stata, and hereafter given the variable name Pw) in the regression model de facto reconstructs the source population while basing the estimation of the corresponding standard error of the association between E and O on the actual observations in (Table 1, rows B and D), rather than on the reconstructed population in (Table 2, row B), the latter being an approach that would underestimate the standard error and thereby overestimate the precision of the RR.

To estimate the RR of outcome O given exposure E using a GLM of the binomial family with a log link and with sampling weight = Pw requires the following command in Stata:

$$glm\ O\ E\ [pweight = Pw],\ family(binomial\ 1)\ link(log)\ eform.$$

eform directs Stata to yield RR instead of ln(RR), which is the default. In our example, the RR is 2.5 with a 95% CI spanning .67 to 9.6. This RR corresponds to the RR point estimate derived from the reconstructed population (Table 2, row B).

Failing to account for the fact that the cases are oversampled, by omitting the sampling weights, as in

$$glm\ O\ E,\ family\ (binomial)\ link(log)\ eform$$

depicted graphically in Figure 3*A*, yields an RR identical to that derived from the exposure-outcome distribution in the combination of the CoCa and the CoCo (Table 1, row H), ie, a substantial overestimation.

Regression analyses carry several other benefits, including the ease of adjusting estimates of associations between E and O for both categorical and continuous confounders. By including an interaction term, they enable us to effectively identify and estimate the size and statistical precision of any effect measure modification between E and D on O. An interaction means that RR_{CoCa} and RR_{CoCo} are different; ie, that getting D changes the risk of getting O given E. Simply adjusting for D would under such circumstances not only violate regression model assumptions, but also iron out any differential effect of E on O between those getting D and the NC.

When estimating the effect of E on O in the underlying cohort, one should refrain from adjusting for D, so that the resulting RR incorporates any effect of E on O mediated through D as well as any interaction between E and D on O. There are, however, situations where adjustment for D is warranted. For instance, to advise public health action, it may be important to break down the effect of E on O by the extent to which it is mediated through D. The size of such mediation can be measured as the relative change in the RR associated with E when estimated from models including and excluding, respectively, D as a covariate. The change in RR of O given E observed by including D as a covariate, ie,

$$glm\ O\ E\ D\ [pweight = Pw],\ family(binomial)\ link(log)\ eform,$$

corresponding to Jiang et al's third ad hoc approach [8] but now with an appropriate balance between cases and NC, would describe the relative change in O risk given E *above and beyond* that mediated by D. In our example, this adjusted RR is 1.97 (95% CI, .49–8.0). The mediation is accordingly 1.97/2.5 = 0.78.

If the 2 models were run on independent data sets, the estimated log RR values for E could be compared using their asymptotic standard errors and their independence. In the present case, however, the 2 models are run on the same data and the 2 estimates of log RR are thus dependent. The dependence may be accounted for with either of 2 different approaches. In Stata, the postestimation command *suest* stores individual score values from the weighted maximum (pseudo)likelihood estimation. The score values are then utilized to compute a robust standard error for the difference of the log RR values in the 2 models. The syntax for a log-binomial regression is

$$\begin{aligned}
&glm\ O\ E\ D\ [iweight = Pw],\ family(binomial)\ link(log)\\
&estimates\ store\ M1\\
&glm\ O\ E\ [iweight = Pw],\ family(binomial)\ link(log)\\
&estimates\ store\ M2\\
&suest\ M1\ M2\\
&lincom\ _b[M1_O:E] - _b[M2_O:E],\ eform
\end{aligned}$$

The suest command requires the sampling weights to be used as "importance weights" (iweight) rather than "probability weights" (pweight).

This yields the (same) point estimate of 0.78 for the mediation and provides us with its 95% CI, which spans .64 to .93. Summarizing, one could say that of the RR = 2.5 that describes the effect of E on O, D contributes with 22% (95% CI, 7%–36%).

Alternatively, a bootstrap approach can be followed. For each bootstrap sample from the observed data, both models are fitted and the difference between log RR values is computed. The usual bootstrap standard errors and CIs can then be computed for the difference, and the CI can be converted to a CI for the ratio of the 2 RRs [13, 14].

If a GLM with a log link for the binomial family does not converge, as may be the case when O is common, or we for other reasons wish to describe the association between E and O with an OR using logistic regression, we can replace the log link with a logit link. GLM of the binomial family with an identity link estimates the absolute risk difference rather than the RR. Using this link enables us to model interactions on an additive scale, which may well be more relevant than doing so on a multiplicative scale in studies such as GEMS, which addresses exposures against which public health interventions, such as vaccination, may be warranted [2, 3, 15].

We have so far considered binary E and O variables, but the RPM is also valid for continuous outcomes. Thus, we can model symmetrically distributed continuous variables, such as infant development scores [16] and growth [17] using an identity link combined with a Gaussian distribution:

$$glm\ O\ E\ [pweight = Pw],\ family(gaussian)\ link(identity),$$

which is equivalent to the simpler linear regression command:

$$regress\ O\ E\ [pweight = Pw].$$

The effect estimate describes the change in O associated with E.

The RPM approach can also be used to model the effects of E on a count, such as that captured in an incidence rate or incidence density, using Poisson regression analysis:

$$glm\ O\ E\ [pweight = Pw],\ family(\ poisson)\ link(log)\ eform.$$

or, when there is overdispersion, using a negative binomial distribution:

$$glm\ O\ E\ [pweight = Pw],\ family(nbinomial)\ link(log)\ eform.$$

The effect estimate describes the incidence rate ratio for O where the exposure is E.

Finally, switching from GLM to time-to-event analysis, the Cox proportional hazards model is well adapted to weighted analysis. Time-to-event analysis requires 2 outcome variables, T is the time from recruitment into the CCF to censoring or to the occurrence of O, which here has the value 1 when the event (eg, death) occurs, or 0 if the individual is censored. In Stata, the sampling weights are included when the data is declared to be time-to-event data:

$$stset\ T\ [pweight = Pw],\ failure(O == 1).$$

The hazard ratios for the event where E is the exposure is returned by

$$stcox\ E.$$

As in CC studies, having served as a control in a CCF study does not preclude an individual from later serving as a case or again being recruited as a control for another case [2, 3]. Similarly, having been enrolled as a case should not bar an individual from again being included as a case, nor from later being included as a control.

The presentation hitherto assumes that we have access to an exact sampling weight. The weight is calculated from the incidence risk, which we cannot obtain from the CCF study. In GEMS, the risk of D is estimated using healthcare utilization and attitude surveys (HUAS), which are undertaken every 4–6 months throughout the study [9, 18]. These estimates carry sampling errors, which need to be taken into account when ultimately estimating the effect of E on O in the underlying cohort.

In the Supplementary Appendix, we provide an Excel sheet, "Data," in the workbook "RPMParametersAndTablesAug2012.xlsx," which generates joint probabilities and 2×2 tables describing an imagined source population based on chosen parameters explained in the sheet "Codes." We used it to generate the 2×2 tables presented in the current manuscript. This population (ie, the underlying cohort) has an exposure (E), a case disease (D), and a dichotomous outcome (O), the latter recorded during follow-up. "Data" enables the user to change the underlying probabilities and associations. In cell C30, it produces an R command highlighted in yellow which, using our R function "rpmBootstrap.R," also provided in the Supplementary Appendix, estimates the composite measures of association, ie, the RR describing the effect of E on O in the reconstructed population ("Unadjusted RR"), the effect of E on O above and beyond that mediated by D ("Adjusted RR"), and the proportion of the Unadjusted RR which is mediated by D ("Mediation RR"). These estimates incorporate not only the sampling error of the CCF study but also that of the D incidence risk estimate obtained from an independent survey. This sampling weight is calculated based on the number of

individuals who developed D (Huas.D) and the number of individuals who did not (Huas.NoD). The analysis might be modified in a variety of ways. For example, the effect of E on O might be modeled in terms of an OR in logistic regression; as a dichotomous outcome on an additive scale, using an identity link to measure risk difference (RD); as the numerator of incidence density or rate in Poisson or negative binomial regression; as a continuous variable in linear regression; or as a hazard ratio in Cox regression.

"rpmBootstrap.R" also generates a Stata (test.dta) and a comma-separated values (test.csv) data set, which contain data from the imagined CCF study and which can be used in a weighted GLM regression of the binomial family to estimate the RR, OR, and RD describing the effect of E on a dichotomous O. This approach, described in some detail in this paper, does not, however, incorporate the sampling error of the sampling weight estimate, and should accordingly be used only when this value is known, as when analyzing data from a CCF study nested in a defined cohort, or when surveys used to estimate D incidence risk are of a size that the derived sampling weights can be considered known values.

EFFECTS OF CHANGING THE POPULATION PARAMETERS

To illustrate how a change in parameters that define critical associations in the underlying population influences the observed effect and to guide the reader on how to use the material in the Supplementary Appendix, let us consider the alterations that occur if we change the association between E and D so that RR changes from 5 to 3. This is achieved by changing RR.D.E in cell D7 of the spreadsheet "Data" in the Workbook "RPMParametersAndTablesAug2012.xlsx" accordingly. The reader will in cell Q84 find that the association described by the RR in the reconstructed population between E and O is reduced from 2.5 to 2.3. Moreover, because we in this example keep the exposure prevalence in the population unchanged at 0.05, the incidence of D is reduced accordingly, in this example from 0.0240 to 0.0220. Such an incidence can be obtained in a survey of 273 individuals that identifies 6 new cases of D.

By running the command returned in cell C30 using the function "rpmBootstrap.R" in R and then the command "*glm O E* [*pweight = Pw*], *family(binomial 1) link(log) eform*" on the generated dataset "test.dta," Stata will return not only the RR of 2.3 but also its 95% CI of .48–11.1. This assumes that the incidence risk of 0.0220 is a fixed number, an assumption which is questionable unless the survey has a very large sample size. Encompassing the sampling error of the sampling weight, our R bootstrap run yielded an RR of 2.4 (95% CI, .26–6.9). Adjusting for D reduced the RR to 2.0 (95% CI, .24–6.3) and quantified the mediation to be 0.86 (95% CI, .69–.96), ie, D contributing with 14% (95% CI, 4%–31%) of the effect of E on O.

If, on the other hand we change the association between E and D so that the RR changes from 5 to 10, the incidence increases to 0.0299, which can be obtained by a survey of 276 individuals of which 8 develop D. Under this scenario, cell Q84 in the sheet "Data" returns an RR of 3.1; Stata also yields its 95% CI of 1.2–7.9. Taking the sampling error of the survey-derived incidence estimate into account using rpmBootstrap, R yielded an RR of 3.4 (95% CI, .95–8.1), which was reduced to 1.9 (95% CI, .58–5.3) after adjustment for D; the mediation was 0.57 (95% CI, .42–.82), so according to this analysis, D contributed with 43% (95% CI, 18%–58%) of the effect of E on O.

The Supplementary Data can also be used to illustrate Jiang et al's argument that, if D changes the risk of O, the first ad hoc approach is valid only if incidence risk is 50% or 0.5. An incidence risk of 50% can be achieved by for example changing the population incidence of D for individuals not exposed to E, ie, p.0.D, to 0.41667. It can be seen that in this unrealistic scenario, Jiang's first ad hoc approach (cell Q54) yields an estimate identical to that obtained with the RPM (cell Q84).

DISCUSSION

We have presented a conceptual framework and illustrate analyses of data from CCF studies. If cases and controls are sampled independently of the exposures and a reliable measure of case disease occurrence can be obtained, such studies can with high efficiency estimate the association between the exposure recorded when the individuals are recruited into the CCF study and outcomes captured during follow-up thereafter. CCF studies exploit the condensation of individuals who develop the case disease into the CoCa, and are thereby more efficient than the corresponding cohort studies.

Previous reports have explored the reuse of CC data to estimate the association between exposures and alternative outcomes [6, 8, 19, 20]. While the suggested approaches range from inverse probability weighting to semiparametric marginal and full likelihood models, the key issue of obtaining appropriate sampling weights is hidden from view. Moreover, there is no suggestion of how to incorporate the standard error of the sampling weight into the composite effect measure generated by the proposed analyses. In general, the rarer the case disease and the smaller the surveys, the more extensive is the contribution from the sampling weight estimates to this joint sampling error.

A well-designed CCF study should be planned with the intent of estimating the association between antecedent exposures and outcomes during follow-up of the 2 cohorts. To enable the necessary weighting, such studies will ensure that

appropriate estimates of case disease incidence, and thereby sampling weight, is captured. This poses particular challenges for CCF studies of infectious diseases, of which the GEMS [10]—to our knowledge—is a conspicuous first. Because the incidence of infectious diseases, such as diarrhea, varies over time and often between relatively closely situated locales, this risk in GEMS is estimated using HUAS rounds undertaken periodically during the study [9, 18], and not as a one-time snapshot [21]. The HUAS-based sampling weights are thereby likely to approach a "real-time" representation of the exposures, case disease, and outcomes, thus increasing the validity of the weighted regression analyses. An estimate of incidence risk derived from such survey data pooled over the duration of the study might be used, if estimates for individual survey rounds seem sufficiently similar. When incidence risk estimates from sequential surveys differ substantially and pooling over time accordingly is questionable, our advice is to first estimate the composite estimates, which describe the effect of E on O for each survey round. This may be of particular relevance for studies that describe microbial agents' contribution to infectious disease, where microepidemics can cause substantial monthly, seasonal, and year-to-year variations [22, 23]. When relevant and appropriate, one can thereafter pool the composite effect sizes, thereby ensuring transparency and epidemiological clarity.

This paper deals with single-population CCF studies that do not recruit controls matched to their corresponding cases, when sampling weights may need to be estimated differently (manuscript in preparation). Further, in a pooled analysis across populations (strata), the weights should be based on the relative stratum sizes and the incidence of D within each stratum.

In most CCF studies relatively few children will be enrolled more than once. Even in a cohort study in Guinea-Bissau, where children were followed with weekly stool specimen examination to identify infections with enteropathogens from birth up to 2 years of age, generalized estimating equations or frailty correction to account for between-child differences did not substantially alter point estimates or precision [24–26]. We suggest that if such correction yields no or little effect on point estimates and their precision, it need not be incorporated in the bootstrapping approaches that capture the sampling error of the sampling weight estimates. It is beyond the scope of this paper to describe in further detail how to take into account between-individual differences in the occurrence of exposures and/or outcomes [27, 28].

In this paper, we describe how a CCF study can be analyzed using weighted regression analysis and, using a bootstrap approach, incorporate the sampling error not only of the CCF component but also of the sampling weight derived from concurrently undertaken survey. Using the spreadsheet and an R function supplied in the Supplementary Appendix, we also show how changes in population parameters, exemplified by a change in the association between E and D, will change the association between E and O in the reconstructed population.

It is our contention that if reliable data on disease incidence are captured, thereby allowing sampling weights to be estimated, weighted regression analysis of CCF data can provide a useful, flexible, and effective analytic tool. We hope that by presenting the conceptual framework for CCF study design and guidance for RPM analysis using weighted regression, we will foster collaboration among infectious disease specialists, epidemiologists, and biostatisticians. Such collaboration in conceptualizing, designing, undertaking, analyzing, and interpreting CCF studies will improve the studies and make it more likely that the analyses and results will address issues of relevance to clinical infectious diseases and communicable disease epidemiology. With constraints on financial and human resources to address critical questions of relationships between specific infections and outcomes (such as clinical sequelae, nutritional impact of infection as well as illness- and infection-associated mortality), CCF studies, because of their efficiency, become particularly attractive. The RPM described in this paper provides a basis for estimating relationships in the population between infection with a pathogen (eg, a diarrheal pathogen as detected in GEMS) and consequences of infection over the period of follow-up. With respect to death possibly related to infection with a diarrheal pathogen, for example, CCF and RPM provide a way to go beyond describing case fatality among enrolled cases who are infected with the pathogen of interest to assessment of the association between the pathogen and mortality in the source population. One would anticipate that this addition to the toolbox of analytic epidemiology might also be useful in estimating the impact of interventions that decrease the frequency of a particular infection on specific outcomes (eg, stunting or death). This can help set priorities for choosing among potential interventions aimed at control of infectious diseases encountered by clinicians on the frontline of clinical care.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/cid/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Financial support. H. K. G., L. vdM., H. Sommerfelt, and H. Steinsland were supported by the University of Bergen; the Norwegian Institute of Public Health; and the Medical Research Council, Cape Town, South Africa; as well as the Norwegian Programme for Development, Research and Education (grant number PRO 48/2002). W. C. B., D. N., T. H. F.,

K. K., and M. M. L. were supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Breslow NE. Statistics in epidemiology: the case-control study. J Am Stat Assoc **1996**; 91:14–28.
2. Rothman KJ. Epidemiology: an introduction. 2nd ed. New York, NY: Oxford University Press, **2012**.
3. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, **2008**.
4. Nagelkerke N, Smits J, le Cessie S, van Houwelingen H. Testing goodness-of-fit of the logistic regression model in case-control studies using sample reweighting. Stat Med **2005**; 24:121–30.
5. Richardson DB, Rzehak P, Klenk J, Weiland SK. Analyses of case-control data for additional outcomes. Epidemiology **2007**; 18:441–5.
6. Reilly M, Torrang A, Klint A. Re-use of case-control data for analysis of new outcome variables. Stat Med **2005**; 24:4009–19.
7. Lumley T. Complex surveys: a guide to analysis using R. Hoboken, NJ: John Wiley, **2010**.
8. Jiang Y, Scott AJ, Wild CJ. Secondary analysis of case-control data. Stat Med **2006**; 25:1323–39.
9. Kotloff KK, Blackwelder WC, Nasrin D, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case-control study. Clin Infect Dis **2012**; 55(S4):S232–45.
10. Levine MM, Kotloff KL, Nataro JP, Muhsen K. Impetus and rationale for the Global Enteric Multicenter Study (GEMS). Clin Infect Dis **2012**; 55(S4):S215–24.
11. Lumley T. Survey-weighted generalised linear models. Available at: http://faculty.washington.edu/tlumley/survey/html/svyglm.html and http://CRAN.R-project.org/package=survey. Accessed 2 November 2012.
12. Lumley T. Survey-weighted Cox models. Available at: http://faculty.washington.edu/tlumley/survey/html/svycoxph.htm and http://CRAN.R-project.org/package=survey. Accessed 2 November 2012.
13. Strand TA, Taneja S, Bhandari N, et al. Folate, but not vitamin B-12 status, predicts respiratory morbidity in north Indian children. Am J Clin Nutr **2007**; 86:139–44.
14. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall, **1993**.
15. Vander Weele TJ. On the distinction between interaction and effect modification. Epidemiology **2009**; 20:863–71.
16. Black MM, Matula K. Essentials of Bayley Scales of Infant Development II assessment. New York: Wiley, **2000**.
17. De Onis M, Garza C, Onyango AW, Martorell R. WHO Child Growth Standards. Acta Paediatrica, **2006**; 95(Suppl 450). Available at: http://www.who.int/entity/childgrowth/standards/Acta_95_S450.pdf.
18. Blackwelder WC, Biswas K, Wu Y, et al. Statistical methods in the Global Enteric Multicenter Study. Clin Infect Dis **2012**; 55(S4): S246–53.
19. Lin DY, Zeng D. Proper analysis of secondary phenotype data in case-control association studies. Genet Epidemiol **2009**; 33: 256–65.
20. Schildcrout JS, Rathouz PJ. Longitudinal studies of binary response data following case-control and stratified case-control sampling: design and analysis. Biometrics **2010**; 66:365–73.
21. World Health Organization. Generic protocols for (i) hospital-based surveillance to estimate the burden of rotavirus gastroenteritis in children and (ii) a community-based survey on utilization of health care services for gastroenteritis in children. Geneva, Switzerland: WHO, **2002**.
22. Mathisen M, Strand TA, Sharma BN, et al. RNA viruses in community-acquired childhood pneumonia in semi-urban Nepal; a cross-sectional study. BMC Med **2009**; 7:35.
23. Mathisen M, Strand TA, Sharma BN, et al. Microepidemics of infections with respiratory viruses in young children in Bhaktapur, Nepal 2004–2007. Available at: http://folk.uib.no/mihtr/CHRP/Virus.html. Accessed 25 June 2012.
24. Valentiner-Branth P, Steinsland H, Fischer TK, et al. Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. J Clin Microbiol **2003**; 41:4238–45.
25. Steinsland H, Valentiner-Branth P, Perch M, et al. Enterotoxigenic *Escherichia coli* infections and diarrhea in a cohort of young children in Guinea-Bissau. J Infect Dis **2002**; 186:1740–7.
26. Steinsland H, Valentiner-Branth P, Gjessing HK, Aaby P, Molbak K, Sommerfelt H. Protection from natural infections with enterotoxigenic *Escherichia coli*: longitudinal study. Lancet **2003**; 362:286–91.
27. Aalen O, Borgan Ø, Gjessing S, SpringerLink (Online service). Survival and event history analysis a process point of view. Statistics for biology and health. New York; London: Springer, **2008**: 539 p.
28. Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata. College Station, Texas: Stata Press, **2012**.

SUPPLEMENT ARTICLE

A Systematic Review and Meta-analysis of the Association Between *Giardia lamblia* and Endemic Pediatric Diarrhea in Developing Countries

Khitam Muhsen and Myron M. Levine

Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

We performed a systematic literature review and meta-analysis examining the association between diarrhea in young children in nonindustrialized settings and *Giardia lamblia* infection. Eligible were case/control and longitudinal studies that defined the outcome as acute or persistent (>14 days) diarrhea, adjusted for confounders and lasting for at least 1 year. Data on *G. lamblia* detection (mainly in stools) from diarrhea patients and controls without diarrhea were abstracted. Random effects model meta-analysis obtained pooled odds ratios (ORs) and 95% confidence intervals (CIs). Twelve nonindustrialized-setting acute pediatric diarrhea studies met the meta-analysis inclusion criteria. Random-effects model meta-analysis of combined results (9774 acute diarrhea cases and 8766 controls) yielded a pooled OR of 0.60 (95% CI, .38–.94; $P = .03$), indicating that *G. lamblia* was not associated with acute diarrhea. However, limited data suggest that initial *Giardia* infections in early infancy may be positively associated with diarrhea. Meta-analysis of 5 persistent diarrhea studies showed a pooled OR of 3.18 (95% CI, 1.50–6.76; $P < .001$), positively linking *Giardia* with that syndrome. The well-powered Global Enteric Multicenter Study (GEMS) is prospectively addressing the association between *G. lamblia* infection and diarrhea in children in developing countries.

Giardia lamblia (synonymous with *Giardia duodenalis* and *Giardia intestinalis*) is a unicellular eukaryotic microscopic enteric protozoa [1–4] that has been incriminated as a cause of diarrhea in individuals in both industrialized and developing countries [5–9]. When clinical illness ensues, it ranges from self-limited acute to persistent diarrhea [4, 10, 11], accompanied by malabsorption. The circumstances under which *G. lamblia* constitutes an etiologic agent of acute or persistent diarrheal disease are not well understood, since in other instances it colonizes without causing diarrhea and in yet other conditions it appears actually to protect against certain forms of diarrheal disease [12, 13].

Experimental challenge studies unequivocally document that some strains of *G. lamblia* can cause diarrhea in healthy adult volunteers [14, 15], and convincing epidemiological descriptions of acute gastroenteritis outbreaks also provide evidence that in certain hosts and settings this protozoan causes acute diarrhea [11, 16–23]. Finally, some case/control studies and longitudinal studies that prospectively follow cohorts of children (and occasionally adults) also support the notion that *G. lamblia* infection is associated with acute or persistent diarrhea [6, 7, 24, 25]. On the other hand, many other case/control and prospective cohort studies do not incriminate *G. lamblia* as a cause of diarrhea [26–28]; moreover, several studies suggest that carriage of this protozoan actually protects against diarrhea [12, 13, 29, 30].

Because of this confusing situation with respect to the role of *G. lamblia* as an enteric pathogen and the

Correspondence: Khitam Muhsen, PhD, Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201 (e-mail: kmuhsen@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S271–93

DOI: 10.1093/cid/cis762

ensuing clinical and epidemiologic equipoise, we systematically reviewed the literature and performed a meta-analysis to examine the association between the occurrence of diarrheal disease in young children in developing countries and the presence of *G. lamblia* in their stool samples. We hypothesized that the association linking *G. lamblia* with diarrhea may differ whether one examines the clinical syndrome of acute vs persistent diarrhea; we further hypothesized that the association may be age-dependent. Finally, we reviewed the role of *G. lamblia* as a putative cause of traveler's diarrhea (albeit mainly in adults), anticipating that these data might shed additional light on the circumstances under which *G. lamblia* causes diarrhea.

METHODS

We performed a PubMed literature search (limited to English-language publications of human studies published prior to 1 April 2012) using the terms "*Giardia* and diarrhea," "*Giardia* gastroenteritis," "*Giardia* and travelers' diarrhea," and "etiology of travelers' diarrhea." To detect additional relevant publications, we used the PubMed option of "related articles" and checked the reference lists of the original and review articles.

Exclusion Criteria

Studies conducted in developed countries, case reports, case series, and studies conducted in patients with immunodeficiency or immunocompromise (eg, human immunodeficiency virus, cancer, post–organ transplantation) were excluded. Also excluded were cross-sectional studies and descriptive studies on the prevalence or detection proportion of *G. lamblia* in patients with diarrhea if they did not include a comparison/control group of subjects without diarrhea. If more than one report was published from the same study, only one report was included. The epidemiologic studies were critically reviewed with special emphasis on whether methodological limitations were evident.

Data Abstraction and Tabulation

Data on study design, study population, sampling frame and sample size, methods to detect *G. lamblia*, definition of diarrhea, case ascertainment, results, and matching or adjusting for potential confounders from each study were abstracted onto standardized forms.

Data extracted from the case/control studies included the number of diarrhea patients and control subjects without diarrhea and the number and/or proportion of cases and controls infected with *G. lamblia*. From the cohort studies we abstracted data on the number of stool samples that were obtained during diarrheal episodes, the number of stool samples that were obtained through routine surveillance in the absence of diarrhea, and the number and proportion of diarrheal and nondiarrheal stools that were positive for *G. lamblia*. Alternatively, depending on the design and the analysis in the original study, data were abstracted on the incidence of diarrheal disease in periods that were classified as *G. lamblia* positive or *G. lamblia* negative. Results stratified by age or other variables were abstracted if they were presented in the original articles. From studies that addressed the association between *G. lamblia* genotype and diarrhea, we abstracted data on the number of patients with diarrhea or gastrointestinal symptoms (cases), number of asymptomatic subjects (controls), the number and percentage of cases and controls infected with genotype A, and the number and percentage of the infected cases and controls with genotype B. If data on the rate ratio or odds ratio (OR) and 95% confidence intervals (CIs) or *P* value were presented in the original article, they were abstracted; otherwise we performed the calculations using WinPepi software version 11.15 [31].

Meta-analysis

Meta-analysis was performed to answer the question of whether *G. lamblia* infection is associated with an increased or a decreased risk of endemic diarrheal disease, using data that were generated by case/control or cohort studies. Pooled measurement of association was obtained using the random effects model and forest plots were generated to display summarized results. Heterogeneity among the studies was tested using heterogeneity χ^2 test and I^2 index [32]. Analyses were performed with stratification by the definition of the outcome (acute vs persistent diarrhea). Potential publication bias was assessed using funnel plots with the log OR of each study on the x-axis plotted against its standard error in the y-axis [33]. We also used the Egger regression intercept [34] to detect publication bias and we performed a cumulative meta-analysis (starting with the largest study) to assess the impact of the study size on the direction of the pooled risk estimate. The Comprehensive Meta-Analysis software package (version 2) was used to produce the analyses [35].

Meta-analysis was restricted to studies conducted in developing countries and other resource-poor settings that presented age-specific findings that allowed data abstraction of the results among children. Additional inclusion criteria were demonstration that matching or adjustment for potential confounders (eg, age, sex) was performed, the definition and duration of diarrhea were provided, and the study endured for at least 1 year (to account for seasonality). In the statistical analyses, we used the adjusted effect estimates of each study to obtain a pooled point estimate. However, if no multivariate analysis was conducted, we used the crude risk estimates.

A PRIMER ON *G. LAMBLIA* INFECTION

Because the biology of *G. lamblia* infection has been previously reviewed [2–4, 10, 36–40], only a few salient features are mentioned in this systematic review of the epidemiology.

The Life Cycle of *G. lamblia*

Giardia lamblia, a unicellular eukaryotic flagellated enteric protozoa [1–4] first described by van Leeuwenhoek in 1681 [2–4], occurs as a nonmotile cyst (responsible for transmission) or a motile trophozoite (associated with clinical symptoms) [4, 10]. Low gastric acidity followed by exposure to pancreatic secretions prompts excystation in the proximal small intestine, with 2 trophozoites deriving from each cyst. The trophozoites replicate in the lumen by binary fission and adhere to enterocytes of the proximal small intestine by suction (using their ventral adhesive disk) [1, 3, 4, 10] and by specific receptor-ligand interactions [36], but do not invade the epithelium. Encystation begins in the small intestine upon exposure to bile salts and is promoted by alkaline pH and decreasing cholesterol levels [1, 3, 4, 10, 41]. Both cysts and trophozoites may be excreted, depending on the nature of the stool. *Giardia* cysts survive in the environment for weeks and months [3, 4, 10], especially in cool and moist conditions [5].

Transmission

Giardia lamblia is transmitted via the ingestion of as few as 10 cysts [42]. Much information on the modes of transmission of *Giardia* comes from studies of infection and illness in industrialized-country settings. Waterborne transmission of *G. lamblia* is well documented [17, 20, 21, 43], including through recreational water activities and swimming [44–48]. The low inoculum facilitates person-to-person transmission among family members [18, 22, 49, 50] and subjects in crowded conditions where hygiene practices may be suboptimal (eg, daycare centers) [16, 18, 22]. Foodborne transmission of *G. lamblia* occurs but is uncommon [16, 51, 52]. Sexual transmission has been reported among men who have sex with men [53–56].

Epidemiological studies [57–61] and genotyping studies of *G. lamblia* support the possibility of zoonotic transmission [62, 63] of *G. lamblia* assemblages A and B, genotypes known to infect both humans and other host species; genotypes C to G infect only animals [64].

Giardia Clinical Illness

Analysis of responses of volunteers to ingestion of *G. lamblia* and descriptions of patients with disease consequent to well-described outbreaks attributed to the protozoan show that the main symptoms include diarrhea, abdominal pain, nausea, vomiting, flatulence, anorexia, and fever [4, 10, 11, 14]. In most instances the diarrheal illness is short-lived and self-limited. However, a proportion of individuals develop persistent diarrhea [4, 10, 11, 65], sometimes accompanied by malabsorption of sugars and fat and by weight loss. In both volunteers and outbreak situations, a sizable proportion of the infected subjects are asymptomatic, often exceeding the proportion who manifest clinical illness [65, 66].

THE ASSOCIATION BETWEEN *G. LAMBLIA* AND DIARRHEAL DISEASE

We identified 46 case/control studies and 18 longitudinal studies conducted from the 1970s through 2009 in developing-country and transitional populations that addressed the association between *G. lamblia* and endemic diarrheal disease.

Overview of the Case/Control Studies

Most of the case/control studies addressed the broad etiology or the role of protozoal agents in acute diarrheal disease, with *G. lamblia* being one of multiple enteropathogens looked for in stools [6, 7, 12, 13, 25, 26, 28, 67–103]. Children comprised the target population in the majority of the studies [6, 7, 12, 13, 25, 26, 28, 67, 68, 70–74, 76, 78, 79, 81, 82, 84–86, 88–91, 93–104], although some studies included adults with or without children [69, 75, 77, 80, 83, 87, 92, 105].

Case ascertainment was performed in the community [78, 96], outpatient clinics [6, 25, 28, 69, 73, 76, 81, 82, 91, 97, 102, 103], emergency rooms [83, 90, 95], or hospitals [7, 12, 13, 26, 68, 70–72, 75, 77, 80, 84, 86, 87, 93, 100, 104, 105]. The control subjects without diarrhea were outpatients [6, 13, 25, 28, 67, 69, 72, 73, 76, 80, 81, 89, 91, 95, 97, 101–103] emergency room patients [83], or hospitalized patients [7, 12, 13, 26, 70, 71, 75, 77, 84, 86, 87, 100, 104, 105], but some studies enrolled community controls [68, 78, 82, 93, 96]. Both inpatient and outpatient settings comprised the sampling frame for some studies [67, 74, 79, 85, 92, 98, 101].

Matching (or adjustment for confounders) between cases and controls by age, sex, and other variables was done in only a fraction of the studies [12, 13, 25, 28, 67, 68, 70, 72, 73, 75, 76, 78–80, 82, 83, 85, 86, 88, 90, 91, 93, 96–98, 102, 103, 105]. In others, neither matching nor adjustment for confounding effects in multivariate analyses was performed [6, 7, 26, 71, 74, 77, 81, 84, 87, 89, 92, 94, 95, 99–101, 104].

Some studies proceeded for at least 1 year [6, 7, 12, 13, 25, 28, 68–73, 75, 79, 80, 82, 84, 86, 90, 92–94, 96, 98, 100], while others lasted only a few months. Stool microscopy was the method most often used for detecting *G. lamblia* [6, 7, 12, 13, 25, 26, 28, 67–69, 71–79, 81–88, 90–97, 100–103]. In a few studies enzyme immunoassay [70, 80, 98, 105] or polymerase chain reaction [89] was used, either in addition to microscopy or as the exclusive method, to detect *Giardia* in stools or duodenal aspirates [104].

The outcome variable was "acute diarrhea" in the majority of the studies [26, 28, 68–73, 75, 76, 80–83, 85, 86, 88, 90, 92, 95, 98, 99, 101, 103, 105], but a few included "persistent diarrhea" as well as "acute diarrhea" [7, 25, 84, 97, 102, 104]; some studies presented pooled results of acute and persistent diarrhea. In 14 studies the length of the diarrheal episode was not specified [6, 12, 13, 67, 74, 77–79, 87, 89, 91, 93, 94, 96] or was not clearly defined [100]. A few studies focused on persistent diarrhea as the outcome variable [106–109], defined as diarrhea that continued unabated for >2 weeks; these studies are presented separately. In one study no operational definition was presented and in another study the authors reported on "chronic diarrhea," defined as diarrhea that lasted >4 weeks [104].

Table 1 summarizes salient features of 12 pediatric case/control studies of acute diarrhea [28, 68, 70, 72, 73, 76, 80, 82, 86, 90, 98, 103] and 3 studies of persistent diarrhea [106, 107, 109] in which the authors controlled for potential confounders by matching or adjusting in multivariable analysis. From one study we abstracted data on children only [80].

Acute Diarrhea

Among studies conducted in children in developing countries or other nonindustrialized settings, 6 studies showed no significant difference between cases and controls in the detection rate of *G. lamblia* [28, 73, 76, 82, 90, 103], whereas 6 other studies showed a significantly lower detection rate of *G. lamblia* in stools from patients with acute diarrhea than from controls [68, 70, 72, 80, 86, 98].

Persistent Diarrhea

Five case/control studies examined the association between *G. lamblia* and persistent diarrhea (≥14 days duration); 4 studies were carried out among children [104, 106, 107, 109] and one study enrolled adults [108]. Table 1 presents the pediatric studies that matched cases and controls according to potential confounders. The detection rate of *G. lamblia* was high in subjects with persistent diarrhea (9.8%–45%) and was 2.6- to 5.9-fold higher than in the control group [106, 107, 109].

Overview of the Longitudinal Studies

The salient features of the study designs and the results from 18 longitudinal studies undertaken in developing countries [24, 27, 30, 110–121] or populations in transition [29, 122, 123] are presented in Table 2. Some longitudinal cohort studies addressed the epidemiology and broad etiology of diarrheal disease [24, 110, 112, 113, 115–121], whereas others confined themselves to addressing the etiologic role of *G. lamblia* in association with diarrhea [27, 29, 30, 111, 114, 122, 123]. Children comprised the study target population except for 2 studies, one from Brazil [110] and the other from Egypt [112], that also included adult household members. The follow-up period in most studies was approximately 24 months [24, 29, 110, 112, 113, 116, 118, 120, 122, 123]. In the remaining studies follow-up was approximately 10–12 months [30, 114, 115, 117, 119], 3 years (average 1.5 years) [111], or 4 years (median 23 months) [27].

The analytical approach compared the prevalence of *G. lamblia* in stool samples that were obtained during diarrheal episodes with the prevalence of the parasite in stools obtained from asymptomatic children or in nondiarrheal stools that were obtained on a systematic predetermined basis (routine surveillance). Two studies compared the incidence of diarrhea in "*G. lamblia*–positive periods" with "*G. lamblia*–negative periods" [27, 29]. In 2 studies the incidence of diarrhea was compared between children who were positive for *G. lamblia* and children whose stool was negative for *G. lamblia* [30, 121]. Measurements of association were reported in only a fraction of the studies [27, 29, 30, 120, 121, 123]. Age- or multivariable-adjusted results were presented in 7 studies [27, 29, 30, 115, 117, 120, 121], whereas the rest presented unadjusted data [24, 111–113, 116, 122, 123]. In some studies, children with diarrhea who provided stool samples were matched with asymptomatic children who delivered stools during routine surveillance, for comparison [112, 113, 115, 117].

Many of the studies did not present the duration of diarrhea. Some studies presented a pooled analysis of acute and persistent diarrhea and a few presented separate analyses for acute vs persistent diarrhea (Table 2).

Only 3 of the 18 cohort studies showed a significantly increased risk of diarrhea in subjects infected with *G. lamblia* [24, 111, 114], while 7 studies showed no significant association between *Giardia* and diarrhea [27, 110, 113, 116–118, 121] (Table 2). One study investigated the length of carriage of *Giardia* in relation to the occurrence of diarrhea but found no significant association [122]. Interestingly, 7 cohort studies actually showed a lower risk for diarrhea in relation to the presence of *G. lamblia* in stools [29, 30, 112, 115, 119, 120, 123].

Is the Association Between G. lamblia and Diarrhea Age-Dependent?

We hypothesized that the association between *G. lamblia* and acute pediatric diarrhea among children in developing countries might be age-dependent; that is, the first infections that occur early in life might be associated with clinical diarrhea, whereas *Giardia* infections in older children might be largely asymptomatic (or may even lower the risk of acute diarrhea).

To address this hypothesis in this review, we abstracted data from studies that presented age-stratified results of the association between *G. lamblia* and diarrhea [24, 27–29, 70, 80, 118, 123]. An impediment to successful pursuit of this analysis was the heterogeneity of the age strata used for reporting data in

Table 1. Case/Control Studies on the Association Between *Giardia lamblia* and Diarrhea Among Children in Nonindustrialized Settings[a]

Study & Country	Study Period	Age	Definition of Diarrhea	*Giardia* Detection	No. Cases Sampling Frame	No. Controls Sampling Frame	*G. lamblia*–Positive Cases, %	*G. lamblia*–Positive Controls, %	OR (95% CI)	Matching/ Adjusting
Acute Diarrhea										
Orlandi [90] Brazil	2000–02	<6 y, 84.5% ≤2 y	Acute diarrhea: ≥3 loose stools in 24 h lasting ≥48 h	Microscopy (cysts)	470 ER	407 ER	1.27%	0.98%	1.30 (.31–6.32)	Age, sex, SES
Huilan [82] Multicenter study in Mexico, Pakistan, China, Myanmar, India	1982–85	<3 y, 47%-75% <1 y	Acute diarrhea: an increase in the number or volume of stools that lasted for ≤72 h. Children with a history of blood or mucus in stools & a temperature of ≥38°C also included	Microscopy (trophozoites or cysts)	Total 3640 outpatient	3279 community	3%	3%	1.00 (.70–1.45)	Region, age, sex, SES, ethnicity
Chatterjee [72] India	1982–83	0–14 y, 32.2% <1 y, 37.5% 1–4 y	Acute diarrhea	Microscopy (trophozoites or cysts)	152 hospital	272 health centers	2.6%	Urban: 25.6% Rural: 15%	0.10 (.04–.28) 0.15 (.04–.49)	Age
Mubashir [86] Pakistan	1983–85	<3 y, 73.6% 1–12 mo	Acute diarrhea of <72 h	Microscopy	402 hospital	402 hospital	2%	8.2%	0.23 (.10–.48)	Age, sex, SES, geographic region, ethnicity
Albert [68] Bangladesh	1994	0–5 y, 80% ≤2 y	Acute diarrhea ≥3 stools/ day	Microscopy	814 ICDDR,B	814 community	0.8%	2.9%	0.30 (.12–.68)	Age, neighborhood
Haque [80] Bangladesh[b]	2004–06	All ages: cases 30% 0–12 mo, controls 19% 0–12 mo	Acute diarrhea: ≥3 abnormal stools in 24 h. Dysentery: the presence of red blood cells, macrophages, or pus cells	EIA	1760 ICDDR,B	1145 clinic	4.5%	15.6%	0.26 (.19–.34)	Age, sex, SES
Hoge [103] Nepal	1994	0.5–5 y, mean age cases 19 mo	Acute diarrhea >3 unformed stools/24 h	Microscopy	124 outpatient	103 community	13%	18%	0.65 (.31–1.36)	Age, sex, neighborhood
Echeverria [73] Thailand	1985–86	<5 y, 80% <2 y	Acute diarrhea: ≥3 loose stools in the previous 24 h for <72 h	Microscopy	1230 outpatient	1230 outpatient	2%	1.3%	1.57 (.84–3.02)	Age
Bodhidatta [70] Thailand	2001–02	3 mo to 5 y, 75% <2 y	Admission due to acute diarrhea	EIA	207 hospital	227 hospital	15%	23%	0.58 (.35–.94)	Age
Loening [28] South Africa	1985–86	<6 y, 83% ≤2 y	≥5 stools/day for >1 d & <7 d	Microscopy (trophozoites or cysts)	373 outpatient	371 outpatient	6.4%	5.9%	1.09 (.60–2.00)	Age, clinic

Table 1 continued.

Study & Country	Study Period	Age	Definition of Diarrhea	*Giardia* Detection	No. Cases Sampling Frame	No. Controls Sampling Frame	*G. lamblia*–Positive Cases, %	*G. lamblia*–Positive Controls, %	OR (95% CI)	Matching/ Adjusting
Gascon [76] Tanzania[c]	1997	0–5 y, mean age: cases 1.9 y, controls 1.6 y	Acute diarrhea ≥3 watery/loose stools/24 h	Microscopy (trophozoites or cysts) Trophozoites	103 clinic	206 clinic	14.5%	15.5%	1.06 (.51–2.19) 1.82 (.76–4.34)	Age, sex, no. of alive siblings, distance to water source, & having a latrine at home
Meng [98] Cambodia[c]	2004–06	3 mo to 5 y, mean age: cases 11.4 mo, controls 31.2 mo	Acute diarrhea ≥3 watery/loose stools/24 h with ≥1 other enteric symptom	EIA	569 inpatient & outpatient	568 inpatient & outpatient	8.3%	21.7%	0.63 (.40–.99)	Age, sex, season
Persistent Diarrhea										
Sullivan [107] Gambia[d]	NA	0.5–3 y	>3 loose stools/day persisting for >2 wk	Microscopy	31 outpatient	33 healthy children outpatient	45%	12%	5.97 (1.50–28.20)	Age, sex
Bhandari [106] India[e]	NA	0–36 mo	Persistent diarrhea ≥3 liquid stools in 24 h lasting ≥14 d; acute diarrhea (<14 d).	Microscopy	175 household surveillance	175 healthy children; 175 acute diarrhea patients	20%	4.6% in each group	5.22 (2.40–12.32)	Age, nutritional status
Mukhopadhyay [109] Nepal[f]	1998–2004	<5 y	Persistent diarrhea: ≥3 liquid stools in 24 h lasting ≥14 d; acute diarrhea (<14 d).	Microscopy	253 inpatient, outpatient	100 healthy community controls, 100 acute diarrhea controls	Trophozoites: 9.8% Cysts: 14.2%	Trophozoites: healthy controls 2%; acute diarrhea 0% Cysts: healthy controls 6%; acute diarrhea 4%	Trophozoites: 5.37 (1.29–47.5) Cysts: 2.60 (1.03–7.79)	Nutritional status

Abbreviations: CI, confidence interval; EIA, enzyme immunoassay; ER, emergency room; ICDDR,B, International Centre for Diarrhoeal Disease Research, Bangladesh; OR, odds ratio; SES, socioeconomic status.

[a] The ORs and 95% CIs were calculated using the raw data that were presented in the original manuscripts, except for 2 studies that presented adjusted OR: Gascon et al [76] and Meng et al [98].

[b] From the study of Haque et al [80] we abstracted data only on children ≤5 years of age.

[c] The adjusted ORs that appeared in the manuscript are presented.

[d] Cases were children with chronic diarrhea and malnutrition; data on the healthy control children are presented.

[e] The results are similar when the control group was the healthy children or the patients with acute diarrhea.

[f] OR was calculated while including the healthy control children.

Table 2. Cohort Studies That Addressed the Role of *Giardia lamblia* in Diarrhea

Study & Country	Study Population	Definition of Diarrhea	Surveillance of Diarrhea	Detection of *Giardia*	No. Diarrheal Stools	No. Nondiarrheal Stools	*G. lamblia*–Positive Diarrhea Stools, %	*G. lamblia*–Positive Nondiarrhea Stools, %	OR/RR (95% CI)[a]
Guerrant [110] Brazil	297, household members, all ages	A significant change in bowel habits: decreased consistency or increased frequency. The duration of diarrhea was presented for other pathogens than *Giardia*.	Daily surveillance was conducted by the mother, and by weekly home visits performed by research assistants.	Microscopy	150	32	6.7%	12.5%	0.50 (.13–2.35)
Schorling [113] Brazil[b]	175, age <5 y	An increase in stool frequency or decrease in consistency, which lasted ≥1 d and was separated from another episode by 3 diarrhea-free days. Acute diarrhea <14 d. Persistent diarrhea ≥14 d,	Home visits 3 times/wk	Microscopy	Acute diarrhea: 50, persistent diarrhea, 40	38	Acute diarrhea: 22%, persistent diarrhea:17.5%	13.2%	1.86 (.59–6.45) 1.40 (.39–5.26)
Newman [111] Brazil	157 newborns followed up from birth	≥3 unformed stools in 24 h. Acute diarrhea lasting <14 d. Persistent diarrhea ≥14 d.	Home visits 3 times/wk.	Microscopy	Acute diarrhea: 514, persistent diarrhea: 97	299	Acute diarrhea: 7.6%, persistent diarrhea: 20.6%	7.4%	1.03 (.58–1.84) 3.27 (1.62–6.62)
Black [116] Peru	153 newborns followed from birth	≥1 d with liquid stools totaling 6 for infants <1 mo, 5 for infants aged 1 mo & 4 for older infants. New episode began after 2 free-illness days. The duration of diarrhea was presented for other pathogens than *Giardia*	Thrice-weekly home visits	Microscopy	952	1973	0.7%	0.8%	0.91 (.35–2.18)
Kaminsky [119] Honduras	266, 101 controls, Age <6 y	An increase in the usual number & change in the consistency of stools for ≥1 d. Acute & persistent diarrhea	Twice-weekly visits	Microscopy (troph. or cysts)	848	101	29%	57%	0.50 (.42–.63)
Hollm-Delgado [27] Peru[c]	220 infants followed up from birth to age 35 mo	≥3 liquid/semi liquid stools/d in 2 consecutive days. The duration of diarrhea was not presented.	Daily home visits	Microscopy	3911	16 973	6%	6.2%	0.95 (.79–1.13)
Boeke [121] Colombia[d]	442, age 5–12 y	Maternal reports. The outcome was diarrhea days.	Daily reports in pictorial diaries	Microscopy (cysts)	Positive children: 28	Negative children: 414	4.0	4.7	0.73 (.52–1.02)
Stanton [24] Bangladesh	343, age <6 y	≥ 3 loose stools in 24 h; new episode began after 14 d without diarrhea. Duration of diarrhea was not presented.	Fortnightly maternal interviews	Microscopy (cysts)	225	1006	11%	4%	2.61 (1.53–4.37)
Baqui [117] Bangladesh	705, age <5 y	≥3 liquid/loose or watery stools or at least 1 bloody stool in 24-hours, Acute diarrhea <14 d, persistent diarrhea ≥4 d	Home visits every fourth day.	Microscopy	Acute diarrhea: 161 Persistent diarrhea: 167	165	Acute diarrhea: 0.6% Persistent diarrhea: 1.2%	1.8%	0.33 0.68
Hasan [118] Bangladesh	252 newborns followed from birth for 2 y	≥3 liquid stools in 24 h or any loose stools accompanied with blood in 24 h. Acute diarrhea <2 wk. Persistent diarrhea ≥2 wk. Data for *Giardia* were presented in a pooled analysis of acute & persistent diarrhea.	Twice-weekly home visits	Microscopy (troph.)	1748	5679	13.2%	13%	1.01 (.86–1.19)
Zaki [112] Egypt	2563, household members, all ages	Reports of the family speak person. The duration of diarrhea was not presented	Twice-weekly home visits	Microscopy	3080	703	44.3%	56.0%	0.63 (.53–.74)
Fraser [123] Israel	164 Bedouin newborns followed from birth to age 23 mo	≥3 soft stools in 24 hours. For infants aged <1 mo ≥4 soft stools. The duration of diarrhea was not presented.	Through the local clinics and hospital, and through monthly and weekly maternal interviews.	Microscopy (cysts)	239	730	22.3%	28.5%	0.8 (.7–.9)

Table 2 continued.

Study & Country	Study Population	Definition of Diarrhea	Surveillance of Diarrhea	Detection of *Giardia*	No. Diarrheal Stools	No. Nondiarrheal Stools	*G. lamblia*–Positive Diarrhea Stools, %	*G. lamblia*–Positive Nondiarrhea Stools, %	OR/RR (95% CI)[a]
Bilenko [29] Israel[e]	238 Bedouin newborns followed from birth to age 23 mo	Maternal reports. The duration of diarrhea (acute vs persistent was not presented).	Weekly maternal interviews	EIA	349	8591	16%	23%	0.65 (.47–.91)
Bilenko [29] Israel[f]	238 Bedouin newborns followed from birth to age 23 mo	Maternal reports. The duration of diarrhea (acute vs persistent was not presented).	Weekly maternal interviews	EIA	1453 *Giardia*-positive months	3001 *Giardia*-negative months	6.7%	6.7%	1.09 (.81–1.46)
Molbak [115] Guinea-Bissau	471–755 children	Maternal reports. Data for *Giardia* were presented in a pooled analysis of acute and persistent diarrhea.	Weekly visits	Microscopy (troph. or cysts) Troph.	1219	511	19.1%, 9.3%	25%, 9.8%	0.8 (.6–1.0) 1.1 (.7–1.5)
Chunge [114] Kenya[g]	84 children aged 10–28 mo	Maternal report. The duration of diarrhea was not presented.	Weekly surveillance	Microscopy (troph. or cysts)	1227	537	78.8%	68.6%	1.69 (1.15–2.54)
Veenemans [30] Tanzania[h]	558, age 6–60 mo	Diarrhea: any report by the caretaker or ≥3 stools in 24 h. The duration of diarrhea was not presented.	Health-facility based surveillance	EIA	Positive children: 192	Negative children: 336	Overall: 0.43 Micronutrients: 0.58 No micronutrients: 0.29	0.68 0.63 0.72	0.84 (.64–1.09) 1.04 (.75–1.43) 0.56 (.34–.90)
Valentiner-Branth [120] Guinea-Bissau[i]	200 newborns followed from birth to age 2 y	Maternal report. The duration of episode was not presented.	Weekly home visits	Microscopy	na	na	na	na	0.64 (.46–.89)

Abbreviations: CI, confidence interval; EIA, enzyme immunoassay; na, not available; OR, odds ratio; RR, rate ratio; Troph, trophozoites.

[a] ORs and CIs were calculated using the raw data presented in the original manuscripts for studies that did not provide measurement of association [24, 110–114, 116, 118, 119]. The measurement of association was provided in the study of Baqui et al [117] (OR) Molbak et al [115] (multivariable adjusted OR), Fraser et al [123] (OR), Bilenko et al [29] (age-adjusted Mantel-Haenszel OR), Hollm-Delgado et al [27] (multivariable adjusted RR), Boeke et al [121] (multivariable adjusted incidence RR), Veenemans et al [30] (adjusted hazard ratio), Valentiner-Branth et al [120] (multivariable adjusted OR).

[b] OR was calculated for acute diarrhea and for persistent diarrhea separately, whereas the comparison group was nondiarrhea.

[c] Adjusted RR for the incidence of diarrheal episodes in *G. lamblia*–positive weeks as compared with *G. lamblia*–negative weeks.

[d] In the study of Boeke et al [121], the incidence of diarrhea days was calculated by dividing the total number of diarrhea days by child years of observation in children who were positive and negative for *Giardia*.

[e] This study [29] presented 2 analyses; this analysis reflects the detection rates of *G. lamblia* in diarrhea stools compared with nondiarrheal stools. Please see the second analysis in the next row.

[f] This study [29] presented 2 analyses; this analysis reflects the adjusted RR for the incidence of diarrheal episodes in *G. lamblia*–positive months as compared with *G. lamblia*–negative months. The first analysis is presented in the previous row.

[g] In the study of Chunge et al [114], the results reflect the detection of *G. lamblia* in stools in relation to maternal reports on diarrhea.

[h] In the study of Veenemans et al [30], the incidence of diarrheal episodes was calculated as the number of episodes divided by child-years of follow-up in children who tested positive and negative for *G. lamblia* at baseline. The results reported in this table are for any reported diarrhea.

[i] Valentiner-Branth et al [120], reported the odds ratio of maternal report on diarrhea during weekly home visits in which stool samples were collected if the child had or did not have diarrhea. The OR in this study reflect the odds of diarrhea during infection with *Giardia*.

Table 3. Association Between *Giardia lamblia* Infection and Diarrhea by Age Groups

Study & Country	Age Groups (mo)	No. *Giardia* Positive/No. Diarrhea	No. *Giardia* Positive/No. Controls	OR (95% CI)[a]
Loening [28] South Africa	0–6	1.2% (n = 80)	1.7% (n = 58)	0.72 (.02–27.05)
	7–12	7.1% (n = 113)	4.3% (n = 115)	1.68 (.52–5.78)
	13–24	8.5% (n = 130)	6.5% (n = 124)	1.34 (.51–3.60)
	25–72	8% (n = 50)	10.8% (n = 74)	0.72 (.18–2.53)
Fraser [123] Israel	≤3	4.2%	1.1%	4.1 (1.1–15.3)
	4–6	5.2%	3.2%	1.6 (.6–4.2)
	7–9	8.7%	11.1%	0.8 (.4–1.4)
	10–12	13.4%	23%	0.5 (.3–0.8)
	13–15	31.8%	33.8%	0.9 (.6–1.3)
	16–18	27.8%	35.9%	0.7 (.4–1.1)
	19–21	41.4%	37%	1.2 (.8–1.9)
	22–24	37.9%	36.1%	1.0 (.6–1.9)
Stanton [24] Bangladesh	<12	3/38 (8%)	0/131 (0%)	…
	12–23	4/55 (7%)	12/173 (7%)	1.05 (.28–3.30)
	24–72	17/132 (13%)	32/702 (5%)	3.10 (1.63–5.72)
Hasan [118] Bangladesh	0–5	2.7% (n = 300)	2.6% (n = 1429)	1.03 (.45–2.16)
	6–11	9.2% (n = 532)	7.3% (n = 1382)	1.29 (.89–1.83)
	12–17	16.5% (n = 520)	16.9% (n = 1405)	0.98 (.74–1.28)
	18–23	22.2% (n = 396)	25.1% (n = 1463)	0.85 (.65–1.11)
Haque [80] Bangladesh	0–12 mo	38/1088 (3.5%)	18/485 (3.7%)	0.94 (.53–1.70)
	1–5 y	4/672 (6.1%)	160/660 (24.2%)	0.20 (.14–0.29)
	6–14 y	31/279 (11.1%)	146/457 (31.9%)	0.27 (.17–0.40)
	15–40 y	91/1222 (7.4%)	92/753 (12.2%)	0.58 (.43–0.79)
	>40 y	4/385 (1%)	24/220 (10.9%)	0.09 (.03–0.24)
Bodhidatta [70] Thailand	3–12	5/85 (6%)	10/103 (10%)	0.58 (.17–1.77)
	13–24	12/79 (15%)	28/77 (36%)	0.31 (.14–0.68)
	25–59	14/43 (33%)	15/47 (32%)	1.00 (.42–2.53)
Studies on the incidence of diarrhea in *Giardia*-positive and -negative periods				

Table 3 continued.

Study & Country	Age Groups (mo)	No. *Giardia* Positive/No. Diarrhea	No. *Giardia* Positive/No. Controls	OR (95% CI)[a]
		No. diarrheal episodes/*Giardia*-positive periods	No. diarrheal episodes/ *Giardia*-negative periods	OR/RR (95% CI)
Bilenko [29] Israel[b]	0–6	3/99	100/1565	0.46 (.11–1.53)
	7–12	45/508	80/914	1.01 (.68–1.51)
	13–18	50/846	20/522	1.58 (.90–2.78)
Hollm-Delgado [27] Peru[c]	0–5	10/188	157/4404	1.56 (.7–3.3)
	6–11	25/402	319/4321	0.86 (.6–1.2)
	12–17	70/762	298/3291	1.02 (.8–1.4)
	18–23	49/901	146/2289	1.00 (.7–1.4)
	24–35	81/1658	135/2668	0.94 (.7–1.4)

Abbreviations: CI, confidence interval; OR, odds ratio; RR, rate ratio.

[a] We calculated the ORs and 95% CIs for Bodhidatta et al [70], Hasan et al [118], Stanton et al [24], and Haque et al [80].

[b] Bilenko et al [29] presented data on the number of diarrheal episodes in months in which *G. lamblia* was detected compared with months in which *G. lamblia* was not detected, and presented OR.

[c] Hollm-Delgado et al [27] presented data on diarrheal stools that were positive for *Giardia* among all *Giardia*-positive stools, compared with diarrheal stools that were negative for *Giardia* among all *Giardia*-negative stools, and presented the adjusted RR.

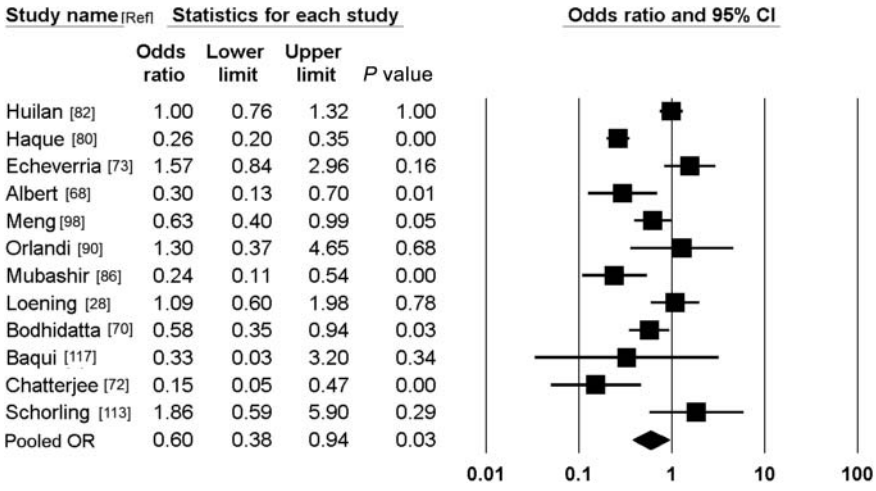

Figure 1. Forest plot of studies on the association between *Giardia lamblia* infection and acute diarrhea among children from developing countries. The odds ratio (OR) and 95% confidence interval (CI) of each study included in the meta-analysis and the pooled OR and 95% CI obtained using the random effects model are presented. Squares and bars represent individual study OR and 95% CI. Diamond represents pooled OR and 95% CI.

the different studies (Table 3). Among these, one study reported a significant increased risk for diarrhea among *Giardia*-infected subjects in the youngest age group (<3 months of age) and a lower risk or no association between *Giardia* and diarrhea in older ages [123] (Table 3).

META-ANALYSIS OF *G. LAMBLIA* AND ENDEMIC PEDIATRIC DIARRHEA

The Association Between *G. lamblia* and Acute Diarrhea

Ten case/control studies [28, 68, 70, 72, 73, 80, 82, 86, 90, 98] and 2 cohort studies [113, 117] that enrolled children from developing countries or other nonindustrialized settings were included in the meta-analysis because their design and execution revealed no fundamental flaws (as explained in the Methods section). From the study of Haque et al [80], we abstracted data only on children aged ≤5 years. These 12 studies [28, 68, 70, 72, 73, 80, 82, 86, 90, 98, 113, 117] fulfilled the inclusion criteria of presenting the outcome variable of acute diarrhea, they matched or controlled for potential confounders, and the study lasted at least 1 year. Using the random effects models, the pooled OR was 0.60 (95% CI, .38–.94; $P = .03$) (Figure 1). This suggests that the presence of *Giardia* infection actually diminished the likelihood of having acute diarrhea among children from developing countries. The heterogeneity test was statistically significant; χ^2 77.9 ($P < .001$), I^2 85.9%.

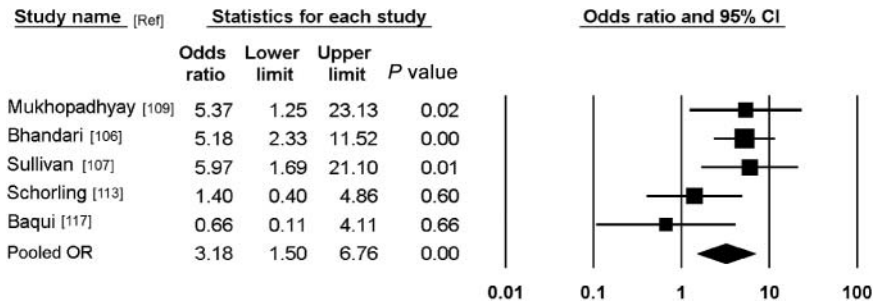

Figure 2. Forest plot of studies on the association between *Giardia lamblia* infection and persistent diarrhea among children from developing countries. The odds ratio (OR) and 95% confidence interval (CI) of each study included in the meta-analysis and the pooled OR and 95% CI obtained using the random effects model are presented. Squares and bars represent individual study OR and 95% CI. Diamond represents pooled OR and 95% CI.

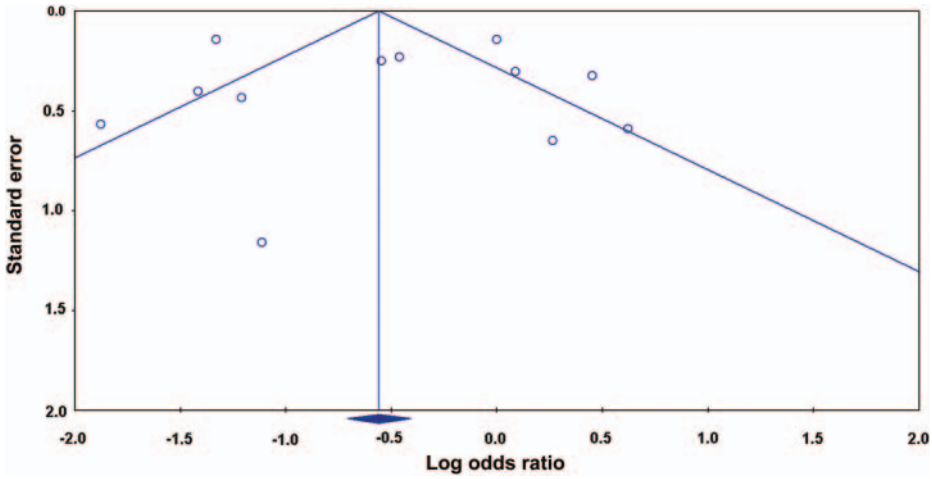

Figure 3. Funnel plot of studies included in the meta-analysis on the association between *Giardia lamblia* infection and acute diarrhea. The log odds ratio (OR) of each study on the x-axis is plotted against the corresponding standard error on the y-axis. The studies are represented in the funnel plot as opened circles. The rhombus shape at the x-axis reflects the log of the pooled OR obtained by using the random effects model.

The Association Between *G. lamblia* and Persistent Diarrhea

Two cohort studies [113, 117] and 3 case/control studies [106, 107, 109] that fulfilled the inclusion criterion of matching between cases and controls and that lasted ≥1 year presented data on persistent diarrhea as an outcome. Results from these 5 studies were combined using the random effects model. The pooled OR was 3.18 (95% CI, 1.50–6.76; $P < .001$), suggesting that *G. lamblia* infection significantly increases the likelihood of persistent diarrhea (Figure 2). The heterogeneity χ^2 test was 7.22 ($P = .125$), I^2 44.6%.

Assessing the Potential of Publication Bias

Figure 3, which presents funnel plots of studies included in the meta-analysis on acute diarrhea, appears visually symmetrical. The Egger regression intercept of the meta-analysis on acute diarrhea studies was 0.32 (95% CI, −3.18 to 3.83; 2-tailed $P = .83$). These results provide no hint of publication bias. Figure 4 shows the cumulative meta-analysis of studies on acute diarrhea. There is no evidence that the addition of the small studies affected the direction of the association between *G. lamblia* infection and the likelihood of acute

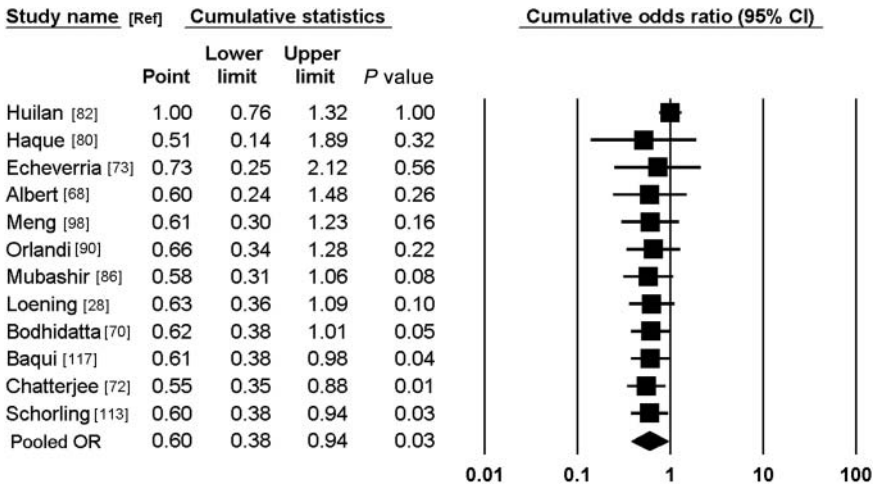

Figure 4. Cumulative meta-analysis of the association between *Giardia lamblia* and acute diarrhea among children from developing countries by study sample size. The change in the pooled odds ratio (OR) is described by adding studies according to their sample size, starting with the largest study. Squares and bars represent individual study OR and 95% CI. Diamond represents pooled OR and 95% CI.

diarrhea in children from developing countries. The Egger regression intercept of the meta-analysis of studies on persistent diarrhea was –2.57 (95% CI, −8.93 to 3.79; 2-tailed $P = .28$).

G. LAMBLIA AND TRAVELERS' DIARRHEA

Early reports of the etiology of travelers' diarrhea [124–126] emphasized enterotoxigenic *Escherichia coli* as the most frequent pathogen, present in approximately 30% of cases [124]. A recent review showed *G. lamblia* in 1.3% and 1.6% of travelers with diarrhea whose destination was Latin America or Africa, respectively, in comparison with 6.2% and 5.7% of travelers to South and Southeast Asia [124]. Only a few studies investigated in parallel the presence of *Giardia* in stools of travelers who did not develop diarrhea (as controls). Our review considered only studies that included a control group of travelers without diarrhea. We identified 12 case/control [127–138] and 2 cohort studies [139, 140] that examined the role of *Giardia* in travelers' diarrhea. Characteristics of these studies are shown in Table 4. *Giardia lamblia* was detected by means of stool microscopy [128, 132–140], except for one study that used enzyme immunoassay [131] and 2 studies that did not specify their method for detecting *G. lamblia* [129, 130].

Studies from the 1970s showed a significant association between travelers' diarrhea and *Giardia* infection among travelers to Leningrad and other sites in the former Soviet Union [133, 139]. Studies of travelers to Nepal revealed that *G. lamblia* was significantly and strongly associated with diarrhea [128, 131], whereas among travelers to Mexico no significant association was found between *G. lamblia* and diarrhea [132, 134, 140]. Three other studies of travelers from Canada, Spain, and the Netherlands (whose destinations were Africa, Asia, and Latin America) also showed a significant positive association between *G. lamblia* and diarrhea or gastrointestinal symptoms [129, 137, 138], and in particular with prolonged (>7 days) and persistent diarrhea (>14 days) [131, 138] (Table 4).

G. LAMBLIA GENOTYPES AND CLINICAL ILLNESS

Experimental challenge studies with *G. lamblia* in which healthy volunteers were inoculated enterally with trophozoites of 2 distinct human isolates of *G. lamblia*, designated GS/M and Isr, demonstrated the establishment of infection and elicitation of clinical illness only among participants who were challenged with the GS/M isolate [14], belonging to *Giardia* assemblage B. This study established the concept that there exists variability among *Giardia* strains with respect to their pathogenicity for humans [14]. Animal experiments support this concept of variable pathogenicity among *Giardia* strains [141].

A few recent studies that were undertaken following the availability of techniques to genotype *G. lamblia* have suggested a possible association between *G. lamblia* genotypes A or B and clinical illness [63, 80, 105, 142–148] (Table 5). Nevertheless, one must exercise caution in drawing conclusions from these preliminary reports as the sample size of each of these studies was small (6–138 diarrheal cases infected with *G. lamblia* and 6–199 nondiarrhea subjects infected with *G. lamblia*). These studies also varied by design, study population, and outcome under investigation (Table 5).

The correlation between *G. lamblia* genotypes and severity of diarrheal illness was examined in an industrialized-country setting among 18 Dutch patients aged 8–60 years with diarrhea and *G. lamblia* infection who visited their general practitioner [149]. Assemblage A *Giardia* was found exclusively among the patients with intermittent/mild disease, while all assemblage B *Giardia* was detected among the patients with more severe cases of diarrhea [149]. Larger studies on the relationship between *G. lamblia* genotype and diarrhea or other gastrointestinal symptoms mostly showed that genotype B was more common (70%–96% in the controls) than genotype A [80, 105, 143, 145, 146, 148, 150], but a higher detection rate of genotype A was found among the symptomatic patients than the controls. A significant association between genotype A and increased risk of diarrhea or other gastrointestinal symptoms was reported from case/control and longitudinal studies [80, 105, 142, 143, 148]. The relationship between *G. lamblia* assemblages and diarrhea was examined in a reanalysis [150] of data from a longitudinal study. This study showed no significant difference between *G. lamblia* genotypes and the number of diarrheal episodes; 0.89 (±0.6) in assemblage A, 1.3 (±1.5) in assemblage B, and 0.80 (±0.84) in mixed infections ($P = .58$) [150]. A study from Sweden compared the distribution of diarrhea and other gastrointestinal symptoms in patients infected with assemblage A *Giardia* (n = 51) and subjects infected with assemblage B (n = 87). The reports on diarrhea were similar between the 2 groups (94% and 99% in assemblage A and B, respectively) but flatulence was more common in subjects infected with assemblage B (85%) than A (65%) [151].

DISCUSSION

The confusing, often conflicting, information in the literature on the role of *G. lamblia* as an enteric pathogen capable of causing diarrheal illness among young children in developing countries led us to undertake this systematic review. Four fundamental conclusions can be drawn from this exercise: (1) *G. lamblia* is capable of causing both acute and persistent

Table 4. Studies That Addressed the Role of Giardia in Travelers' Diarrhea

Study	Country of Origin	Definition of Outcome	Sampling	Age	Travel Destination	*Giardia*-Positive Cases/Total Cases (%)	*Giardia*-Positive Controls/Total Controls (%)	RR/OR (95% CI)[a]	Matching/ Adjusting
Andersson [139]	Sweden	Gastrointestinal symptoms	Students who traveled to Leningrad	Adult students	Leningrad	27/27 (100%)	3/11 (27.3%)	3.66 (1.75–10.26)	None
Brodsky [133]	US	Gastrointestinal symptoms	Tourists, CDC surveillance	All ages	Former Soviet Union	83/153 (54.3%)	8/153 (5.2%)	21.49 (10.11–44.49)	None
Merson [140]	US, Canada, Netherlands, England	The occurrence between 12 h after arrival to Mexico City and 5 d after departure of any unformed stool not attributed to a preexisting condition plus ≥1 enteric symptom. Or ≥3 watery stools in 24 h	Physicians & their family members	Mainly adults	Mexico	1/51 (2%)	1/43 (2.3%)	0.84 (.01–67.49)	None
DuPont [132]	US, Venezuela, Mexico	Acute diarrhea: unformed bowel movements at a daily rate twice of the usual rate plus ≥1 enteric symptom	University clinic	Adult students	Mexico	US 6% (total cases 77), LA 18% (total cases 18)	US 3% (total controls 67), LA 11% (total controls 27)	2.26 (.43–17.20), 1.60 (.19–13.43)	Country of origin, length of stay in Mexico
Bolivar [134]	US, Venezuela, Mexico	Unformed bowel movement at daily rate twice that of the subject's usual rate & ≥1 other enteric symptom	University clinic	Adult students	Mexico	3/91 (3.3%)	2/74 (2.7%)	1.23 (.18–10.54)	Country of origin, length of stay in Mexico
Back [127]	Sweden	≥2 abnormal loose stools/d	Swedish battalion in United Nations forces	Adults	Cyprus	1/79 (1.3%)	0/66 (0%)		Serving conditions (next bedfellow)
Echeverria [135]	US	≥3 loose stools or ≥2 loose stools with other enteric symptom	Soldiers who attended a clinic	Adults	Philippines	3/152 (2%)	2/58 (3.5%)	0.56 (.06–6.94)	None
Hoge[b] [136]	Foreign residents & tourists	Change in normal bowel movements with ≥3 loose stools in 24 h	CIWEC, USEM	All ages	Nepal	7/148 (4.7%)	1/95 (1%)	4.67 (.58–212.52)	Group matching by clinic & season
Shlim [128]	Tourists, expatriates	Change in normal bowel movements & ≥3 loose stools in 24 h	CIWEC	≥18 y	Nepal	25/189 (13.2%)	3/112 (2.6%)	5.54 (1.62–29.23)	None
Gascon [137]	Spain	Diarrhea that occurred between 12 h after arriving in & 5 d after departing from the travel country. Diarrhea ≥3 watery stools in 24 h, or unformed stools plus enteric symptom	Tropical Medicine Department	NA	Asia, Africa, Central & Latin America	11/165 (6.7%)	3/165 (1.8%)	3.86 (.99–21.86)	Area visited, controls were relatives or travel companions of cases

Table 4 continued.

Study	Country of Origin	Definition of Outcome	Sampling	Age	Travel Destination	*Giardia*-Positive Cases/Total Cases (%)	*Giardia*-Positive Controls/Total Controls (%)	RR/OR (95% CI)[a]	Matching/ Adjusting
Schultsz [138]	Netherlands	≥3 loose stools in 24 h, any number of watery stools in 24 h, or 1–2 loose stools in 24 h plus ≥1 enteric symptom	Outpatient Department for Tropical Diseases	2–75 y	Asia, Africa, Central & Latin America	Acute 2% (total cases 49), persistent 16.4% (total cases 116)	4.9% (total controls 102)	0.40 (.01–3.78), 3.80 (1.30–13.48)	None
Boggild [129]	Canada	Diagnosis of giardiasis	Tropical Disease Unit. (GeoSentinel Network)	Mean 37.3 y	International travel	69/1622 (4.3%)	5/1906 (0.3%)	16.9 (6.8–41.9)	None
Paschke [130]	Germany	≥3 unformed stools in 24 h plus ≥1 symptom of enteric infection	Department of Infectious Diseases & Tropical Medicine.	2–80 y	Asia, Latin America, Europe, other	7/114 (6.1%)	3/56 (5.4%)	1.16 (.25–7.20)	None
Pandey [131]	US, Japan, Australia, New Zealand, Western Europe	≥3 unformed stools in 24 h	CIWEC	>18 y	Nepal	Overall 42/372 (11.3%) ≤7 d 7%, >7 d 26%	5 (2.9%)	3.75 (1.40–9.98), 2.48 (.95–7.52), 11.78 (4.41–35.90)	Age, sex, nationality, resident/ tourist status, length of stay in Nepal, season

Abbreviations: CDC, Centers for Disease Control and Prevention; CI, confidence interval; LA, Latin American (Venezuela and Mexico); CIWEC, Canadian International Water and Energy Consultants; OR, odds ratio; RR, rate ratio; USEM, US Embassy Medical Care.

[a] ORs and 95% CIs were calculated using the abstracted data from each study. For the study of Andersson et al [139], RR was calculated. Boggild et al [129] reported crude OR and Pandey et al [131] reported adjusted OR.

[b] Data from the study of Hoge et al [136] were abstracted on 148 cases of diarrhea among which coccidian-like organisms were not identified. Cases and controls from both clinics (CIWEC and USEM) were pooled.

Table 5. Association Between *Giardia lamblia* Assemblage and Diarrhea or Other Gastrointestinal Symptoms

Study & Country	Design	Subjects	Outcome	No. Cases Genotyped	No. Controls Genotyped	Genotype A Cases, No. (%)	Genotype A Controls, No. (%)	Genotype B Cases, No. (%)	Genotype B Controls, No. (%)	OR (95% CI)[a]
Paintlia [147] India	Case series	Adults from gastroenterology & dermatology clinics	Gastrointestinal symptoms: diarrhea, weight loss, abdominal pain	6	6	4 (66.7%)	1 (16.7%)	2 (33.3%)	5 (83.3%)	10.0 (.43–588.32)
Eligio-Garcia [63] Mexico	Case series	6–12 y old children	Chronic/recurrent diarrhea & abdominal pain	6	7	6 (100%)	7 (100%)	0 (0%)	0 (0%)	
Al-Mohammed [144][b] Saudi Arabia	Cross-sectional	Primary school-age children 6–12 y	Acute & chronic diarrhea	24	16	7 (29.2%)	16 (100%)	15 (62.5%)	0 (0%)	
Molina [146][c] Argentina	Cross-sectional	2–14 y old enrolled at health centers or public schools	Symptoms: diarrhea, anorexia, vomiting, abdominal pain	50	41	8 (16%)	6 (14.6%)	42 (84%)	35 (85.4%)	1.11 (.30–4.28)
Aydin [143] Turkey	Case/control	Patients from Dept of Infectious Disease & Gastroenterology	Diarrhea	20	24	17 (85%)	2 (8%)	3 (15%)	22 (92%)	62.33 (9.13–480.26)
Sahagun [148][d] Spain	Case/control	Ages 2–72 y from outpatient clinic suspected of parasitosis	Symptoms: diarrhea, nausea, abdominal pain/cramps, weight loss, flatulence	55	49	29 (52.7%)	14 (28.5%)	26 (47.3%)	35 (71.5%)	2.79 (1.23–6.38)
Haque [105][e] Bangladesh	Matched case/control	All ages, cases from ICDDR,B, Hospital controls	Diarrhea	84	199	16 (19.5%)	20 (10.5%)	68 (80.5%)	179 (89.5%)	2.11 (1.04–4.26)
Haque [80][f] Bangladesh	Matched case/control	All ages, cases from ICDDR,B, Clinic controls	Acute diarrhea	138	184	29 (21%)	10 (5.4%)	109 (79%)	174 (94.6%)	4.63 (2.20–10.27)
Read [142] Australia	Longitudinal	Children in day care centers age <5 y	Diarrhea	9	14	6 (66.7%)	1 (7.1%)	3 (33.3%)	13 (92.9%)	26.0 (2.2–304.7)

Table 5 continued.

Study & Country	Design	Subjects	Outcome	No. Cases Genotyped	No. Controls Genotyped	Genotype A Cases, No. (%)	Genotype A Controls, No. (%)	Genotype B Cases, No. (%)	Genotype B Controls, No. (%)	OR (95% CI)[a]
Ajjampur [145][g] India	Longitudinal	Newborns followed till age 3 y	Acute & intermediate diarrhea (<14 d)	45	50	5 (11.1%)	2 (4%)	40 (89.9%)	48 (96%)	3.00 (.46–32.74)

Abbreviations: CI, confidence interval; ICDDR,B, International Centre for Diarrhoeal Disease Research, Bangladesh; OR, odds ratio.

[a] The OR presented here reflects the odds of *G. lamblia* genotype A infection among the cases in comparison to odds of genotype A infection in the control group. The calculations of OR (95% CI) were made using the raw data in the original manuscripts when the authors did not present the measurement of association [80, 143–145, 148].

[b] Two samples with mixed infections among the cases were not included in the calculation.

[c] Samples with mixed infections (n = 3) were not included in the calculation. Cases were children with gastrointestinal symptoms.

[d] Four samples had mixed A and B genotypes, 2 among the symptomatic and 2 among the asymptomatic patients [148]; they were not included in the data presented in this table. Among genotype A isolates, only subgenotype AII was identified [148].

[e] A total of 267 *G. lamblia*–positive stool specimens were genotyped, among which 16 samples harbored mixed A and B genotypes that were counted twice by the authors, once as A genotype and once as B genotype [105].

[f] *G. lamblia*–positive stools of 144 and 199 cases and controls were genotyped; of these 6 and 15 were mixed genotype A and B infections [80], and they were not included in the calculations presented in this table. Part of the *G. lamblia* genotypes included in this study was reported in an earlier report [105].

[g] Five mixed infections among the cases and 1 in the control group were excluded from the analysis.

diarrheal illness in adult and pediatric hosts who reside in industrialized countries, including following exposure when they travel to developing countries. (2) *G. lamblia* does not generally cause acute pediatric diarrhea among infants and children in developing countries, although limited data suggest that infants in the first trimester of life may experience acute clinical diarrhea in response to presumed initial *G. lamblia* infections. (3) *G. lamblia* is positively associated with persistent diarrhea among children in developing countries. (4) Genotyping suggests that 2 *G. lamblia* genotypes (assemblages A and B) may be particularly pathogenic for humans.

Among residents of industrialized countries, evidence from experimental challenge studies of adult volunteers [14, 15], investigations of (particularly water-borne) outbreaks of diarrheal disease [11, 16–18, 23], and investigations of travelers who visit developing countries or known endemic areas [128, 129, 133, 137] collectively and convincingly document that *G. lamblia* can cause acute diarrheal illness and other gastrointestinal disease. In contrast, as summarized in this review, contradictory results have been reported from epidemiological studies performed in subjects residing in developing countries [24, 72, 82].

Herein we provide the first systematic review and meta-analysis that attempts to address the etiologic role of *G. lamblia* in relation to diarrheal illness among children from developing countries or other nonindustrialized settings where *Giardia* is highly endemic. In systematically reviewing the literature, it became apparent that among the many published studies that explored a possible association between *G. lamblia* and diarrhea, few utilized rigorous design methodology and analytical techniques. For example, few studies controlled for potential confounders and many lacked the statistical power to detect differences between patients with diarrhea and controls without diarrhea. There were very few birth cohort studies, thus the age of first infection could not be assessed. Some studies did not differentiate between the clinical syndromes of acute vs persistent diarrhea, which is critical for analyzing data on *Giardia* infections; consequently, misclassification of the outcome variable may have ensued. Finally, some studies were limited in duration, covering <1 year.

Accordingly, we limited our analysis to the case/control and cohort studies that utilized rigorous methodology and controlled for potential confounders, lasted ≥ 1 year, and clearly defined the outcome variable (ie, acute vs persistent diarrhea). In so doing, we found there was no significant association between the presence of *Giardia* in stools and increased risk of acute diarrhea among children living developing countries or nonindustrialized settings [28, 73, 82, 90, 113, 117]. Indeed, there was evidence of a significant inverse association between the presence of *Giardia* in stools and acute diarrhea among children in developing country or other nonindustrialized

settings [68, 70, 72, 80, 86, 98]. A pooled analysis of the studies that utilized rigorous methodology showed that *G. lamblia* was associated with a 40% lower likelihood of acute diarrhea in children from developing countries ($P = .03$) (Figure 1).

One may invoke differences in the host, the parasite, or host-parasite interactions to explain the strikingly distinct responses to *Giardia* exposure among children and adults from industrialized countries vs developing countries. The former are at risk of developing acute diarrhea when they encounter *G. lamblia*, whereas pediatric subjects in the latter settings experience apparent innocuity or even a protective effect of *Giardia* against acute diarrhea when infected with this protozoan. One possible explanation may relate to the age of initial exposure and the frequency of subsequent reexposure. In developing-country populations, *G. lamblia* is ubiquitous and the initial infection is acquired in the first few weeks of life [27, 66, 118, 120, 123, 152, 153]. In developing-country settings, the initial or first few *G. lamblia* infections may result in diarrhea [74, 123] but immunity is rapidly acquired, thereupon conferring protection against symptomatic disease when subsequently exposed. *Giardia lamblia* gastroenteritis outbreaks in daycare centers in Canada provide indirect support for this explanation [18, 154]. Children of Canadian origin and those from other industrialized countries were more likely to be infected and to develop *Giardia* illness compared with children of immigrant families from developing countries [18, 154].

One well-established mechanism by which infants and young children in developing countries are protected against symptomatic disease upon exposure to *Giardia* is by suckling on mothers whose breast milk contains high titers of anti-*Giardia* secretory immunoglobulin A (SIgA). Breastfeeding is strongly associated with protection against clinical *Giardia* diarrhea, even though it does not generally prevent acquisition of *G. lamblia* infection or chronic carriage [152, 155]. Importantly, clinical protection is correlated with levels of specific anti-*G. lamblia* SIgA in milk [156]. Analogous evidence derives from experimental challenges of adult US volunteers [14]. Secretory IgA anti-*Giardia* antibodies were detected in duodenal fluids of subjects who experienced diarrhea following initial challenge with *Giardia* strain Gsm and these SIgA antibodies correlated with protection against clinical disease when the subjects were rechallenged but not with prevention of reinfection. There are also reports of anti-*Giardia* properties of breast milk due to moieties other than specific SIgA [157, 158]. Breast milk–derived passive protection may allow the child to acquire active immunity upon exposure to *G. lamblia* without paying the price of a clinically overt initial infection.

Another possible explanation for the apparent divergent clinical responses to *Giardia* in industrialized vs developing country pediatric populations may reside in differences in the small intestine. Young children in industrialized countries harbor low numbers of bacteria in their proximal small intestine and their mucosal architecture is characterized by elongated villi and modest numbers of intraepithelial and lamina propria lymphocytes. In contrast, the "normal" small intestine of young children living in impoverished, fecally contaminated conditions in developing countries is marked by blunted villi and hypercellularity of the lamina propria and by small bowel bacterial overgrowth [159–162]. While there is a spectrum of severity of such changes, they are collectively referred to as "environmental enteropathy" (or "tropical enteropathy") [159–162]. When the small intestine of the young child in the industrialized country setting is exposed to *G. lamblia,* acute diarrhea or other symptomatology not uncommonly results. In contrast, among young children in developing countries who often manifest environmental enteropathy, *G. lamblia* appears more often to result in asymptomatic colonization without acute diarrhea. In the environmental enteropathy gut, *G. lamblia* may modulate the innate immune system and mucosal environment such that a degree of protection is conferred against diarrhea caused by other enteropathogens. In vitro studies show that intestinal mucus may affect *Giardia* activity [163], and studies in mice suggest that the normal gut flora may play a role in susceptibility to *Giardia* infection [164]. If this phenomenon is also true in humans, it is possible that these factors might affect the clinical presentation of *Giardia* infection.

Whereas our systematic review did not find an association between *G. lamblia* infection and increased risk of acute diarrhea in children in developing-country settings, *Giardia* was significantly associated with persistent diarrhea in these pediatric populations [106, 107, 109]. The clinical illness of patients with giardiasis in industrialized settings who were infected during outbreaks or during travel to endemic areas also shows that symptoms may persist for several weeks [21, 154, 165, 166].

One must ponder why *G. lamblia* appears to be associated with a 3-fold increase in the risk of persistent diarrhea among children in developing countries but the pathogen is not associated with an increased risk of acute diarrhea. One hypothesis is that the infants and young children who develop persistent *G. lamblia* diarrhea constitute a subset of high-risk pediatric hosts because they have more severe chronic undernutrition than their peers of the same age (usually manifest as severe stunting), more severe environmental enteropathy or due to a genetic predisposition (such as combined IgA and immunoglobulin G2 deficiency).

Attributes of the parasite may also account for the propensity to cause persistent diarrhea in children in developing countries. Preliminary evidence supports an association between *G. lamblia* genotype A and B in the development of clinically overt diarrhea and other gastrointestinal symptoms

[80, 105, 142, 143, 148]. Antigenic variation manifested by *Giardia* may also play a role in the outcome or course of infection. *Giardia lamblia* trophozoites have variant specific proteins that coat the entire parasite including its flagella. Trophozoites can switch these proteins every 6.5–13.5 generations and this may allow evasion of the immune response, establishment of more persistent infection [36, 167–171], and a propensity to persistent diarrhea.

This systematic review supports the contention that asymptomatic *G. lamblia* infection somehow protects against diarrheal illness, although mechanistically it is not obvious how this occurs. *Giardia lamblia* infection triggers both host innate [36–38] and adaptive immune responses [36–38, 172]. Secretion of innate antimicrobial products having anti-*Giardia* activity (eg, defensin, lactoferrin) by the intestinal epithelium [36–38] and nitric oxide and reactive oxygen species has been described [37, 38]. Secretion of mucins [36, 37] and glycoproteins of the intestinal mucus layer can reduce attachment of a broad range of pathogens to the mucosal surface [36]. These responses elicited by *Giardia* may negatively affect other pathogens in the gut. Thus, repetitive or prolonged *Giardia* trophozoite attachment to the intestinal epithelium for extended periods may render the mucosa unfavorable for the attachment of other enteropathogens. *Giardia lamblia* has also been shown to bind cholera enterotoxin [173] and heavy *Giardia muris* infection significantly diminishes the intestinal secretion stimulated by cholera toxin compared to mouse intestine without *Giardia* [174]. Thus, *Giardia* may offer protection against otherwise severe diarrhea caused by enterotoxigenic bacterial pathogens like *Vibrio cholerae* and enterotoxigenic *E. coli*. Finally, one report suggests that the severity of rotavirus gastroenteritis in Bedouin infants may have been significantly reduced in the presence of *Giardia lamblia* coinfection [29].

Giardia lamblia infection induces serum immunoglobulin M and intestinal SIgA anti-*Giardia* antibodies [14, 36–38, 172], of which local SIgA is considered the most important for controlling and clearing the infection [37, 38]. Interleukin 6 and T-dependent responses have also been described [36–38]. These anti-*Giardia* responses may contribute to nonspecific or cross-protection against other enteropathogens [152].

In summary, evidence does not incriminate *G. lamblia* as a cause of acute diarrhea in young children in developing countries but does suggest an important role of *G. lamblia* infection in persistent diarrhea in such populations. Statistically well-powered, controlled studies such as the Global Enteric Multicenter Study (GEMS) are needed to clarify the circumstances under which *G. lamblia* infection may be involved in the development of diarrheal disease. In 7 developing-country sites, GEMS will help address whether *Giardia* infections in early infancy are positively linked to moderate-to-severe diarrhea, whether some pediatric hosts (eg, more stunted) are more prone to develop persistent diarrhea, whether *Giardia* decreases the risk of acute diarrhea from other specific enteropathogens, and whether specific *Giardia* genotypes exhibit enhanced pathogenicity over other genotypes.

Notes

Financial support. This study was funded by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Marquardt WC, Demaree RS, Grieve RB. Giardia and giardiosis. Parasitology and vector biology. 2nd ed. San Diego, CA: Academic Press, **2000**:89–96.
2. Adam RD. Biology of *Giardia lamblia*. Clin Microbiol Rev **2001**; 14:447–75.
3. Ankarklev J, Jerlstrom-Hultqvist J, Ringqvist E, Troell K, Svard SG. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat Rev Microbiol **2010**; 8:413–22.
4. Farthing MJ. Giardiasis. Gastroenterol Clin North Am **1996**; 25:493–515.
5. Lane S, Lloyd D. Current trends in research into the waterborne parasite *Giardia*. Crit Rev Microbiol **2002**; 28:123–47.
6. Nakano T, Binka FN, Afari EA, et al. Survey of enteropathogenic agents in children with and without diarrhoea in Ghana. J Trop Med Hyg **1990**; 93:408–12.
7. Nimri LF, Meqdam M. Enteropathogens associated with cases of gastroenteritis in a rural population in Jordan. Clin Microbiol Infect **2004**; 10:634–9.
8. Yoder JS, Harral C, Beach MJ. Giardiasis surveillance—United States, 2006–2008. MMWR Surveill Summ **2010**; 59:15–25.
9. European Centre for Disease Prevention and Control. Annual epidemiological report on communicable diseases in Europe 2010. European Center for Disease Prevention and Control: Stockholm, Sweden, **2010**.
10. Lebwohl B, Deckelbaum RJ, Green PH. Giardiasis. Gastrointest Endosc **2003**; 57:906–13.
11. Katz DE, Heisey-Grove D, Beach M, Dicker RC, Matyas BT. Prolonged outbreak of giardiasis with two modes of transmission. Epidemiol Infect **2006**; 134:935–41.
12. al-Bwardy MA, Ramia S, al-Frayh AR, et al. Bacterial, parasitic and viral enteropathogens associated with diarrhoea in Saudi children. Ann Trop Paediatr **1988**; 8:26–30.
13. Khan MM, Iqbal J, Ghafoor A, Burney MI. Aetiologic agents of diarrhoeal diseases in hospitalised children in Rawalpindi, Pakistan. J Diarrhoeal Dis Res **1988**; 6:228–31.
14. Nash TE, Herrington DA, Losonsky GA, Levine MM. Experimental human infections with *Giardia lamblia*. J Infect Dis **1987**; 156:974–84.
15. Rendtorff RC, Holt CJ. The experimental transmission of human intestinal protozoan parasites. IV. Attempts to transmit *Endamoeba coli* and *Giardia lamblia* cysts by water. Am J Hyg **1954**; 60:327–38.
16. White KE, Hedberg CW, Edmonson LM, Jones DB, Osterholm MT, MacDonald KL. An outbreak of giardiasis in a nursing home with evidence for multiple modes of transmission. J Infect Dis **1989**; 160:298–304.

17. Shaw PK, Brodsky RE, Lyman DO, et al. A communitywide outbreak of giardiasis with evidence of transmission by a municipal water supply. Ann Intern Med **1977**; 87:426–32.
18. Keystone JS, Krajden S, Warren MR. Person-to-person transmission of *Giardia lamblia* in day-care nurseries. Can Med Assoc J **1978**; 119:241–2, 247–8.
19. Addiss DG, Davis JP, Roberts JM, Mast EE. Epidemiology of giardiasis in Wisconsin: increasing incidence of reported cases and unexplained seasonal trends. Am J Trop Med Hyg **1992**; 47:13–9.
20. Lopez CE, Dykes AC, Juranek DD, et al. Waterborne giardiasis: a communitywide outbreak of disease and a high rate of asymptomatic infection. Am J Epidemiol **1980**; 112:495–507.
21. Birkhead G, Vogt RL. Epidemiologic surveillance for endemic *Giardia lamblia* infection in Vermont. The roles of waterborne and person-to-person transmission. Am J Epidemiol **1989**; 129:762–8.
22. Black RE, Dykes AC, Sinclair SP, Wells JG. Giardiasis in day-care centers: evidence of person-to-person transmission. Pediatrics **1977**; 60:486–91.
23. Nygard K, Schimmer B, Sobstad O, et al. A large community outbreak of waterborne giardiasis-delayed detection in a non-endemic urban area. BMC Public Health **2006**; 6:141.
24. Stanton B, Silimperi DR, Khatun K, et al. Parasitic, bacterial and viral pathogens isolated from diarrhoeal and routine stool specimens of urban Bangladeshi children. J Trop Med Hyg **1989**; 92:46–55.
25. Nimri LF, Elnasser Z, Batchoun R. Polymicrobial infections in children with diarrhoea in a rural area of Jordan. FEMS Immunol Med Microbiol **2004**; 42:255–9.
26. Shetty N, Narasimha M, Raghuveer TS, Elliott E, Farthing MJ, Macaden R. Intestinal amoebiasis and giardiasis in southern Indian infants and children. Trans R Soc Trop Med Hyg **1990**; 84:382–4.
27. Hollm-Delgado MG, Gilman RH, Bern C, et al. Lack of an adverse effect of *Giardia intestinalis* infection on the health of Peruvian children. Am J Epidemiol **2008**; 168:647–55.
28. Loening WE, Coovadia YM, van den Ende J. Aetiological factors of infantile diarrhoea: a community-based study. Ann Trop Paediatr **1989**; 9:248–55.
29. Bilenko N, Levy A, Dagan R, Deckelbaum RJ, El-On Y, Fraser D. Does co-infection with *Giardia lamblia* modulate the clinical characteristics of enteric infections in young children? Eur J Epidemiol **2004**; 19:877–83.
30. Veenemans J, Mank T, Ottenhof M, et al. Protection against diarrhea associated with *Giardia intestinalis* is lost with multi-nutrient supplementation: a study in Tanzanian children. PLoS Negl Trop Dis **2011**; 5:e1158.
31. Abramson JH. WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. Epidemiol Perspect Innov **2004**; 1:6.
32. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med **2002**; 21:1539–58.
33. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol **2001**; 54:1046–55.
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ **1997**; 315:629–34.
35. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive meta-analysis version 2. Englewood, NJ: Biostat, **2005**.
36. Muller N, von Allmen N. Recent insights into the mucosal reactions associated with *Giardia lamblia* infections. Int J Parasitol **2005**; 35:1339–47.
37. Roxstrom-Lindquist K, Palm D, Reiner D, Ringqvist E, Svard SG. Giardia immunity—an update. Trends Parasitol **2006**; 22:26–31.
38. Eckmann L. Mucosal defences against *Giardia*. Parasite Immunol **2003**; 25:259–70.
39. Buret AG. Pathophysiology of enteric infections with *Giardia duodenalis*. Parasite **2008**; 15:261–5.
40. Cotton JA, Beatty JK, Buret AG. Host parasite interactions and pathophysiology in *Giardia* infections. Int J Parasitol **2011**; 41:925–33.
41. Gillin FD, Reiner DS, Boucher SE. Small-intestinal factors promote encystation of *Giardia lamblia* in vitro. Infect Immun **1988**; 56:705–7.
42. Rendtorff RC. The experimental transmission of human intestinal protozoan parasites. II. *Giardia lamblia* cysts given in capsules. Am J Hyg **1954**; 59:209–20.
43. Weniger BG, Blaser MJ, Gedrose J, Lippy EC, Juranek DD. An outbreak of waterborne giardiasis associated with heavy water runoff due to warm weather and volcanic ashfall. Am J Public Health **1983**; 73:868–72.
44. Isaac-Renton JL, Philion JJ. Factors associated with acquiring giardiasis in British Columbia residents. Can J Public Health **1992**; 83:155–8.
45. Hoque ME, Hope VT, Kjellstrom T, Scragg R, Lay-Yee R. Risk of giardiasis in Aucklanders: a case-control study. Int J Infect Dis **2002**; 6:191–7.
46. Hoque ME, Hope VT, Scragg R, Kjellstrom T. Children at risk of giardiasis in Auckland: a case-control analysis. Epidemiol Infect **2003**; 131:655–62.
47. Gray SF, Gunnell DJ, Peters TJ. Risk factors for giardiasis: a case-control study in Avon and Somerset. Epidemiol Infect **1994**; 113:95–102.
48. Dennis DT, Smith RP, Welch JJ, et al. Endemic giardiasis in New Hampshire: a case-control study of environmental risks. J Infect Dis **1993**; 167:1391–5.
49. Balcioglu IC, Limoncu E, Ertan P, Yereli K, Ozbilgin A, Onag A. Incidence of giardiasis among siblings in Turkey. Pediatr Int **2003**; 45:311–3.
50. Chute CG, Smith RP, Baron JA. Risk factors for endemic giardiasis. Am J Public Health **1987**; 77:585–7.
51. Porter JD, Gaffney C, Heymann D, Parkin W. Food-borne outbreak of *Giardia lamblia*. Am J Public Health **1990**; 80:1259–60.
52. Mintz ED, Hudson-Wragg M, Mshar P, Cartter ML, Hadler JL. Foodborne giardiasis in a corporate office setting. J Infect Dis **1993**; 167:250–3.
53. Meyers JD, Kuharic HA, Holmes KK. *Giardia lamblia* infection in homosexual men. Br J Vener Dis **1977**; 53:54–5.
54. Keystone JS, Keystone DL, Proctor EM. Intestinal parasitic infections in homosexual men: prevalence, symptoms and factors in transmission. Can Med Assoc J **1980**; 123:512–4.
55. Phillips SC, Mildvan D, William DC, Gelb AM, White MC. Sexual transmission of enteric protozoa and helminths in a venereal-disease-clinic population. N Engl J Med **1981**; 305:603–6.
56. Pakianathan MR, McMillan A. Intestinal protozoa in homosexual men in Edinburgh. Int J STD AIDS **1999**; 10:780–4.
57. Pereira MG, Atwill ER, Barbosa AP. Prevalence and associated risk factors for *Giardia lamblia* infection among children hospitalized for diarrhea in Goiania, Goias State, Brazil. Rev Inst Med Trop Sao Paulo **2007**; 49:139–45.
58. Coles CL, Levy A, Dagan R, Deckelbaum RJ, Fraser D. Risk factors for the initial symptomatic *Giardia* infection in a cohort of young Arab-Bedouin children. Ann Trop Paediatr **2009**; 29:291–300.
59. Mahmud MA, Chappell C, Hossain MM, Habib M, Dupont HL. Risk-factors for development of first symptomatic *Giardia* infection among infants of a birth cohort in rural Egypt. Am J Trop Med Hyg **1995**; 53:84–8.
60. Boontanom P, Mungthin M, Tan-Ariya P, Naaglor T, Leelayoova S. Epidemiology of giardiasis and genotypic characterization of *Giardia duodenalis* in preschool children of a rural community, central Thailand. Trop Biomed **2011**; 28:32–9.
61. Ratanapo S, Mungthin M, Soontrapa S, et al. Multiple modes of transmission of giardiasis in primary schoolchildren of a rural community, Thailand. Am J Trop Med Hyg **2008**; 78:611–5.
62. Sprong H, Caccio SM, van der Giessen JW. Identification of zoonotic genotypes of *Giardia duodenalis*. PLoS Negl Trop Dis **2009**; 3:e558.

63. Eligio-Garcia L, Cortes-Campos A, Jimenez-Cardoso E. Genotype of *Giardia* intestinalis isolates from children and dogs and its relationship to host origin. Parasitol Res **2005**; 97:1–6.
64. Monis PT, Andrews RH, Mayrhofer G, Ey PL. Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. Infect Genet Evol **2003**; 3:29–38.
65. Moore G, Cross W, McGuire D, et al. Epidemic giardiasis at a ski resort. N Engl J Med **1969**; 281:402–27.
66. Flanagan PA. Giardia—diagnosis, clinical course and epidemiology. A review. Epidemiol Infect **1992**; 109:1–22.
67. Addy PA, Antepim G, Frimpong EH. Prevalence of pathogenic *Escherichia coli* and parasites in infants with diarrhoea in Kumasi, Ghana. East Afr Med J **2004**; 81:353–7.
68. Albert MJ, Faruque AS, Faruque SM, Sack RB, Mahalanabis D. Case-control study of enteropathogens associated with childhood diarrhea in Dhaka, Bangladesh. J Clin Microbiol **1999**; 37:3458–64.
69. Al-Gallas N, Bahri O, Bouratbeen A, Ben Haasen A, Ben Aissa R. Etiology of acute diarrhea in children and adults in Tunis, Tunisia, with emphasis on diarrheagenic *Escherichia coli*: prevalence, phenotyping, and molecular epidemiology. Am J Trop Med Hyg **2007**; 77:571–82.
70. Bodhidatta L, McDaniel P, Sornsakrin S, Srijan A, Serichantalergs O, Mason CJ. Case-control study of diarrheal disease etiology in a remote rural area in Western Thailand. Am J Trop Med Hyg **2010**; 83:1106–9.
71. Casalino M, Yusuf MW, Nicoletti M, et al. A two-year study of enteric infections associated with diarrhoeal diseases in children in urban Somalia. Trans R Soc Trop Med Hyg **1988**; 82:637–41.
72. Chatterjee BD, Thawani G, Sanyal SN. Etiology of acute childhood diarrhoea in Calcutta. Trop Gastroenterol **1989**; 10:158–66.
73. Echeverria P, Taylor DN, Lexsomboon U, et al. Case-control study of endemic diarrheal disease in Thai children. J Infect Dis **1989**; 159:543–8.
74. El-Hakim MA, El-Sahn A. Association of parasites and diarrhoea among children less than five years of age in a rural area in Egypt. J Egypt Public Health Assoc **1996**; 71:439–63.
75. Gambhir IS, Jaiswal JP, Nath G. Significance of *Cryptosporidium* as an aetiology of acute infectious diarrhoea in elderly Indians. Trop Med Int Health **2003**; 8:415–9.
76. Gascon J, Vargas M, Schellenberg D, et al. Diarrhea in children under 5 years of age from Ifakara, Tanzania: a case-control study. J Clin Microbiol **2000**; 38:4459–62.
77. Gassama A, Sow PS, Fall F, et al. Ordinary and opportunistic enteropathogens associated with diarrhea in Senegalese adults in relation to human immunodeficiency virus serostatus. Int J Infect Dis **2001**; 5:192–8.
78. Georges MC, Roure C, Tauxe RV, et al. Diarrheal morbidity and mortality in children in the Central African Republic. Am J Trop Med Hyg **1987**; 36:598–602.
79. Gracey M, Burke V, Robinson J. Patterns of intestinal infection in Australian Aboriginal children. Ann Trop Paediatr **1983**; 3:35–9.
80. Haque R, Mondal D, Karim A, et al. Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. Clin Infect Dis **2009**; 48:1191–7.
81. Hassan EM, el-Meneza SA, el-Rashidy Z, Rashad R, Rabie S, Fahmy SA. Detection of enteropathogens in diarrhoeal diseases among malnourished Egyptian infant and children. J Egypt Public Health Assoc **1989**; 64:461–74.
82. Huilan S, Zhen LG, Mathan MM, et al. Etiology of acute diarrhoea among children in developing countries: a multicentre study in five countries. Bull World Health Organ **1991**; 69:549–55.
83. Korzeniowski OM, Dantas W, Trabulsi LR, Guerrant RL. A controlled study of endemic sporadic diarrhoea among adult residents of southern Brazil. Trans R Soc Trop Med Hyg **1984**; 78:363–9.
84. Menon BS, Abdullah S, Mahamud F, et al. Low prevalence of *Cryptosporidium* parvum in hospitalized children in Kota Bharu, Malaysia. Southeast Asian J Trop Med Public Health **2001**; 32:319–22.
85. Mertens TE, Wijenayake R, Pinto MR, et al. Microbiological agents associated with childhood diarrhoea in the dry zone of Sri Lanka. Trop Med Parasitol **1990**; 41:115–20.
86. Mubashir M, Khan A, Baqai R, et al. Causative agents of acute diarrhoea in the first 3 years of life: hospital-based study. J Gastroenterol Hepatol **1990**; 5:264–70.
87. Nahrevanian H, Assmar M, Samin MG. Cryptosporidiosis among immunocompetent patients with gastroenteritis in Iran: a comparison with other enteropathogenic parasites. J Microbiol Immunol Infect **2007**; 40:154–6.
88. Ogunsanya TI, Rotimi VO, Adenuga A. A study of the aetiological agents of childhood diarrhoea in Lagos, Nigeria. J Med Microbiol **1994**; 40:10–4.
89. Opintan JA, Newman MJ, Ayeh-Kumi PF, et al. Pediatric diarrhea in southern Ghana: etiology and association with intestinal inflammation and malnutrition. Am J Trop Med Hyg **2010**; 83:936–43.
90. Orlandi PP, Magalhaes GF, Matos NB, et al. Etiology of diarrheal infections in children of Porto Velho (Rondonia, Western Amazon region, Brazil). Braz J Med Biol Res **2006**; 39:507–17.
91. Orlandi PP, Silva T, Magalhaes GF, et al. Enteropathogens associated with diarrheal disease in infants of poor urban areas of Porto Velho, Rondonia: a preliminary study. Mem Inst Oswaldo Cruz **2001**; 96:621–5.
92. Oyofo BA, Subekti D, Tjaniadi P, et al. Enteropathogens associated with acute diarrhea in community and hospital patients in Jakarta, Indonesia. FEMS Immunol Med Microbiol **2002**; 34:139–46.
93. Paniagua GL, Monroy E, Garcia-Gonzalez O, Alonso J, Negrete E, Vaca S. Two or more enteropathogens are associated with diarrhoea in Mexican children. Ann Clin Microbiol Antimicrob **2007**; 6:17.
94. Sallon S, el-Shawwa R, Khalil M, et al. Diarrhoeal disease in children in Gaza. Ann Trop Med Parasitol **1994**; 88:175–82.
95. Sanchez-Vega JT, Tay-Zavala J, Aguilar-Chiu A, et al. Cryptosporidiosis and other intestinal protozoan infections in children less than one year of age in Mexico City. Am J Trop Med Hyg **2006**; 75:1095–8.
96. Simango C, Dindiwe J. The aetiology of diarrhoea in a farming community in Zimbabwe. Trans R Soc Trop Med Hyg **1987**; 81: 552–3.
97. Ming ZF, Xi ZD, Dong CS, et al. Diarrhoeal disease in children less than one year of age at a children's hospital in Guangzhou, People's Republic of China. Trans R Soc Trop Med Hyg **1991**; 85:667–9.
98. Meng CY, Smith BL, Bodhidatta L, et al. Etiology of diarrhea in young children and patterns of antibiotic resistance in Cambodia. Pediatr Infect Dis J **2011**; 30:331–5.
99. Househam KC, Dove MG, Smith MS. Enteropathogens associated with acute infantile diarrhoea in Bloemfontein, South Africa. J Trop Pediatr **1987**; 33:287–8.
100. Ajjampur SS, Rajendran P, Ramani S, et al. Closing the diarrhoea diagnostic gap in Indian children by the application of molecular techniques. J Med Microbiol **2008**; 57(Pt 11):1364–8.
101. Na'was TE, Abo-Shehada MN. A study of the bacterial and parasitic causes of acute diarrhoea in northern Jordan. J Diarrhoeal Dis Res **1991**; 9:305–9.
102. Stintzing G, Mollby R, Habte D. Enterotoxigenic *Escherichia coli* and other enteropathogens in paediatric diarrhoea in Addis Ababa. Acta Paediatr Scand **1982**; 71:279–86.
103. Hoge CW, Echeverria P, Rajah R, et al. Prevalence of *Cyclospora* species and other enteric pathogens among children less than 5 years of age in Nepal. J Clin Microbiol **1995**; 33:3058–60.
104. Rajeshwari K, Jaggi N, Aggarwal V, Kalra KK, Mittal SK, Baveja U. Determinants of symptomatic giardiasis in childhood. Trop Gastroenterol **1996**; 17:70–6.

105. Haque R, Roy S, Kabir M, Stroup SE, Mondal D, Houpt ER. Giardia assemblage A infection and diarrhea in Bangladesh. J Infect Dis **2005**; 192:2171–3.
106. Bhandari N, Bahl R, Dua T, Kumar R, Srivastava R. Role of protozoa as risk factors for persistent diarrhea. Indian J Pediatr **1999**; 66:21–6.
107. Sullivan PB, Marsh MN, Phillips MB, et al. Prevalence and treatment of giardiasis in chronic diarrhoea and malnutrition. Arch Dis Child **1991**; 66:304–6.
108. Yakoob J, Abbas Z, Beg MA, et al. Prevalences of *Giardia lamblia* and *Cryptosporidium parvum* infection in adults presenting with chronic diarrhoea. Ann Trop Med Parasitol **2010**; 104:505–10.
109. Mukhopadhyay C, Wilson G, Pradhan D, Shivananda PG. Intestinal protozoal infestation profile in persistent diarrhea in children below age 5 years in western Nepal. Southeast Asian J Trop Med Public Health **2007**; 38:13–9.
110. Guerrant RL, Kirchhoff LV, Shields DS, et al. Prospective study of diarrheal illnesses in northeastern Brazil: patterns of disease, nutritional impact, etiologies, and risk factors. J Infect Dis **1983**; 148:986–97.
111. Newman RD, Moore SR, Lima AA, Nataro JP, Guerrant RL, Sears CL. A longitudinal study of *Giardia lamblia* infection in north-east Brazilian children. Trop Med Int Health **2001**; 6:624–34.
112. Zaki AM, DuPont HL, el Alamy MA, et al. The detection of enteropathogens in acute diarrhea in a family cohort population in rural Egypt. Am J Trop Med Hyg **1986**; 35:1013–22.
113. Schorling JB, Wanke CA, Schorling SK, McAuliffe JF, de Souza MA, Guerrant RL. A prospective study of persistent diarrhea among children in an urban Brazilian slum. Patterns of occurrence and etiologic agents. Am J Epidemiol **1990**; 132:144–56.
114. Chunge RN, Nagelkerke N, Karumba PN, et al. Longitudinal study of young children in Kenya: intestinal parasitic infection with special reference to *Giardia lamblia*, its prevalence, incidence and duration, and its association with diarrhoea and with other parasites. Acta Trop **1991**; 50:39–49.
115. Molbak K, Wested N, Hojlyng N, et al. The etiology of early childhood diarrhea: a community study from Guinea-Bissau. J Infect Dis **1994**; 169:581–7.
116. Black RE, Lopez de Romana G, Brown KH, Bravo N, Bazalar OG, Kanashiro HC. Incidence and etiology of infantile diarrhea and major routes of transmission in Huascar, Peru. Am J Epidemiol **1989**; 129:785–99.
117. Baqui AH, Sack RB, Black RE, et al. Enteropathogens associated with acute and persistent diarrhea in Bangladeshi children less than 5 years of age. J Infect Dis **1992**; 166:792–6.
118. Hasan KZ, Pathela P, Alam K, et al. Aetiology of diarrhoea in a birth cohort of children aged 0–2 year(s) in rural Mirzapur, Bangladesh. J Health Popul Nutr **2006**; 24:25–35.
119. Kaminsky RG. Parasitism and diarrhoea in children from two rural communities and marginal barrio in Honduras. Trans R Soc Trop Med Hyg **1991**; 85:70–3.
120. Valentiner-Branth P, Steinsland H, Fischer TK, et al. Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. J Clin Microbiol **2003**; 41:4238–45.
121. Boeke CE, Mora-Plazas M, Forero Y, Villamor E. Intestinal protozoan infections in relation to nutritional status and gastrointestinal morbidity in Colombian school children. J Trop Pediatr **2010**; 56:299–306.
122. Fraser D, Bilenko N, Deckelbaum RJ, Dagan R, El-On J, Naggan L. *Giardia lamblia* carriage in Israeli Bedouin infants: risk factors and consequences. Clin Infect Dis **2000**; 30:419–24.
123. Fraser D, Dagan R, Naggan L, et al. Natural history of *Giardia lamblia* and *Cryptosporidium* infections in a cohort of Israeli Bedouin infants: a study of a population in transition. Am J Trop Med Hyg **1997**; 57:544–9.
124. Shah N, DuPont HL, Ramsey DJ. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. Am J Trop Med Hyg **2009**; 80:609–14.
125. Riddle MS, Sanders JW, Putnam SD, Tribble DR. Incidence, etiology, and impact of diarrhea among long-term travelers (US military and similar populations): a systematic review. Am J Trop Med Hyg **2006**; 74:891–900.
126. Black RE. Pathogens that cause travelers' diarrhea in Latin America and Africa. Rev Infect Dis **1986**; 8(suppl 2):S131–5.
127. Back E, Jonsson M, Wadstrom T. Enterotoxin-producing bacteria stools from Swedish United Nations soldiers in Cyprus. Infection **1978**; 6:116–20.
128. Shlim DR, Hoge CW, Rajah R, Rabold JG, Echeverria P. Is *Blastocystis hominis* a cause of diarrhea in travelers? A prospective controlled study in Nepal. Clin Infect Dis **1995**; 21:97–101.
129. Boggild AK, Yohanna S, Keystone JS, Kain KC. Prospective analysis of parasitic infections in Canadian travelers and immigrants. J Travel Med **2006**; 13:138–44.
130. Paschke C, Apelt N, Fleischmann E, et al. Controlled study on enteropathogens in travellers returning from the tropics with and without diarrhoea. Clin Microbiol Infect **2011**; 17:1194–200.
131. Pandey P, Bodhidatta L, Lewis M, et al. Travelers' diarrhea in Nepal: an update on the pathogens and antibiotic resistance. J Travel Med **2011**; 18:102–8.
132. DuPont HL, Olarte J, Evans DG, Pickering LK, Galindo E, Evans DJ. Comparative susceptibility of Latin American and United States students to enteric pathogens. N Engl J Med **1976**; 295:1520–1.
133. Brodsky RE, Spencer HC Jr, Schultz MG. Giardiasis in American travelers to the Soviet Union. J Infect Dis **1974**; 130:319–23.
134. Bolivar R, Conklin RH, Vollet JJ, et al. Rotavirus in travelers' diarrhea: study of an adult student population in Mexico. J Infect Dis **1978**; 137:324–7.
135. Echeverria P, Blacklow NR, Zipkin C, et al. Etiology of gastroenteritis among Americans living in the Philippines. Am J Epidemiol **1979**; 109:493–501.
136. Hoge CW, Shlim DR, Rajah R, et al. Epidemiology of diarrhoeal illness associated with coccidian-like organism among travellers and foreign residents in Nepal. Lancet **1993**; 341:1175–9.
137. Gascon J, Vargas M, Quinto L, Corachan M, Jimenez de Anta MT, Vila J. Enteroaggregative *Escherichia coli* strains as a cause of traveler's diarrhea: a case-control study. J Infect Dis **1998**; 177:1409–12.
138. Schultsz C, van den Ende J, Cobelens F, et al. Diarrheagenic *Escherichia coli* and acute and persistent diarrhea in returned travelers. J Clin Microbiol **2000**; 38:3550–4.
139. Andersson T, Forssell J, Sterner G. Outbreak of giardiasis: effect of a new antiflagellate drug, tinidazole. Br Med J **1972**; 2:449–51.
140. Merson MH, Morris GK, Sack DA, et al. Travelers' diarrhea in Mexico. A prospective study of physicians and family members attending a congress. N Engl J Med **1976**; 294:1299–305.
141. Cevallos A, Carnaby S, James M, Farthing JG. Small intestinal injury in a neonatal rat model of giardiasis is strain dependent. Gastroenterology **1995**; 109:766–73.
142. Read C, Walters J, Robertson ID, Thompson RC. Correlation between genotype of *Giardia duodenalis* and diarrhoea. Int J Parasitol **2002**; 32:229–31.
143. Aydin AF, Besirbellioglu BA, Avci IY, Tanyuksel M, Araz E, Pahsa A. Classification of *Giardia duodenalis* parasites in Turkey into groups A and B using restriction fragment length polymorphism. Diagn Microbiol Infect Dis **2004**; 50:147–51.
144. Al-Mohammed HI. Genotypes of *Giardia intestinalis* clinical isolates of gastrointestinal symptomatic and asymptomatic Saudi children. Parasitol Res **2011**; 108:1375–81.
145. Ajjampur SS, Sankaran P, Kannan A, et al. *Giardia duodenalis* assemblages associated with diarrhea in children in South India identified by PCR-RFLP. Am J Trop Med Hyg **2009**; 80:16–9.

146. Molina N, Minvielle M, Grenovero S, Salomon C, Basualdo J. High prevalences of infection with *Giardia intestinalis* genotype B among children in urban and rural areas of Argentina. Ann Trop Med Parasitol **2011**; 105:299–309.
147. Paintlia AS, Descoteaux S, Spencer B, et al. *Giardia lamblia* groups A and B among young adults in India. Clin Infect Dis **1998**; 26:190–1.
148. Sahagun J, Clavel A, Goni P, et al. Correlation between the presence of symptoms and the *Giardia duodenalis* genotype. Eur J Clin Microbiol Infect Dis **2008**; 27:81–3.
149. Homan WL, Mank TG. Human giardiasis: genotype linked differences in clinical symptomatology. Int J Parasitol **2001**; 31:822–6.
150. Kohli A, Bushen OY, Pinkerton RC, et al. *Giardia duodenalis* assemblage, clinical presentation and markers of intestinal inflammation in Brazilian children. Trans R Soc Trop Med Hyg **2008**; 102:718–25.
151. Lebbad M, Petersson I, Karlsson L, et al. Multilocus genotyping of human *Giardia* isolates suggests limited zoonotic transmission and association between assemblage B and flatulence in children. PLoS Negl Trop Dis **2011**; 5:e1262.
152. Mahmud MA, Chappell CL, Hossain MM, Huang DB, Habib M, DuPont HL. Impact of breast-feeding on *Giardia lamblia* infections in Bilbeis, Egypt. Am J Trop Med Hyg **2001**; 65:257–60.
153. Miotti PG, Gilman RH, Santosham M, Ryder RW, Yolken RH. Age-related rate of seropositivity of antibody to *Giardia lamblia* in four diverse populations. J Clin Microbiol **1986**; 24:972–5.
154. Wright RA, Spencer HC, Brodsky RE, Vernon TM. Giardiasis in Colorado: an epidemiologic study. Am J Epidemiol **1977**; 105:330–6.
155. Morrow AL, Reves RR, West MS, Guerrero ML, Ruiz-Palacios GM, Pickering LK. Protection against infection with *Giardia lamblia* by breast-feeding in a cohort of Mexican infants. J Pediatr **1992**; 121:363–70.
156. Walterspiel JN, Morrow AL, Guerrero ML, Ruiz-Palacios GM, Pickering LK. Secretory anti-*Giardia lamblia* antibodies in human milk: protective effect against diarrhea. Pediatrics **1994**; 93:28–31.
157. Gillin FD, Reiner DS, Wang CS. Human milk kills parasitic intestinal protozoa. Science **1983**; 221:1290–2.
158. Hernell O, Ward H, Blackberg L, Pereira ME. Killing of *Giardia lamblia* by human milk lipases: an effect mediated by lipolysis of milk lipids. J Infect Dis **1986**; 153:715–20.
159. Neto UF, Martins MCD, Lima FLD, Patricio FRS, Toledo MRF. Asymptomatic environmental enteropathy among slum-dwelling infants. J Am Coll Nutr **1994**; 13:51–6.
160. Anderson CM, Langford RF. Bacterial content of small intestine of children in health, in coeliac disease, and in fibrocystic disease of pancreas. Br Med J **1958**; 1:803–6.
161. Fagundes-Neto U, Viaro T, Wehba J, Patricio FR, Machado NL. Tropical enteropathy (environmental enteropathy) in early childhood: a syndrome caused by contaminated environment. J Trop Pediatr **1984**; 30:204–9.
162. dos Reis JC, de Morais MB, Oliva CA, Fagundes-Neto U. Breath hydrogen test in the diagnosis of environmental enteropathy in children living in an urban slum. Dig Dis Sci **2007**; 52:1253–8.
163. Zenian A, Gillin FD. Interactions of *Giardia lamblia* with human intestinal mucus: enhancement of trophozoite attachment to glass. J Protozool **1985**; 32:664–8.
164. Singer SM, Nash TE. The role of normal flora in *Giardia lamblia* infections in mice. J Infect Dis **2000**; 181:1510–2.
165. Hanevik K, Hausken T, Morken MH, et al. Persisting symptoms and duodenal inflammation related to *Giardia duodenalis* infection. J Infect **2007**; 55:524–30.
166. Kent GP, Greenspan JR, Herndon JL, et al. Epidemic giardiasis caused by a contaminated public water supply. Am J Public Health **1988**; 78:139–43.
167. Prucca CG, Rivero FD, Lujan HD. Regulation of antigenic variation in *Giardia lamblia*. Annu Rev Microbiol **2011**; 65:611–30.
168. Nash TE, Herrington DA, Levine MM, Conrad JT, Merritt JW Jr. Antigenic variation of *Giardia lamblia* in experimental human infections. J Immunol **1990**; 144:4362–9.
169. Nash TE. Surface antigenic variation in *Giardia lamblia*. Mol Microbiol **2002**; 45:585–90.
170. Nash TE, Lujan HT, Mowatt MR, Conrad JT. Variant-specific surface protein switching in *Giardia lamblia*. Infect Immun **2001**; 69:1922–3.
171. Singer SM, Elmendorf HG, Conrad JT, Nash TE. Biological selection of variant-specific surface proteins in *Giardia lamblia*. J Infect Dis **2001**; 183:119–24.
172. Ali SA, Hill DR. *Giardia intestinalis*. Curr Opin Infect Dis **2003**; 16:453–60.
173. McCardell BA, Madden JM, Stanfield JT, Tall BD, Stephens MJ. Binding of cholera toxin to *Giardia lamblia*. J Clin Microbiol **1987**; 25:1786–8.
174. Ljungstrom I, Holmgren J, Svennerholm AM, Ferrante A. Changes in intestinal fluid and mucosal immune responses to cholera toxin in *Giardia muris* infection and binding of cholera toxin to *Giardia muris* trophozoites. Infect Immun **1985**; 50:243–9.

SUPPLEMENT ARTICLE

Diagnostic Microbiologic Methods in the GEMS-1 Case/Control Study

Sandra Panchalingam,[1] Martin Antonio,[2] Anowar Hossain,[3] Inacio Mandomando,[4] Ben Ochieng,[5] Joseph Oundo,[5] T. Ramamurthy,[6] Boubou Tamboura,[7] Anita K. M. Zaidi,[8] William Petri,[9] Eric Houpt,[9] Patrick Murray,[10] Valeria Prado,[11] Roberto Vidal,[11] Duncan Steele,[12] Nancy Strockbine,[10] Philippe Sansonetti,[13] Roger I. Glass,[10] Roy M. Robins-Browne,[14] Marija Tauschek,[14] Ann-Marie Svennerholm,[15] Karen Kotloff,[1] Myron M. Levine,[1] and James P. Nataro[1,16]

[1]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; [2]Medical Research Council, Fajara, The Gambia; [3]International Center for Diarrhoeal Disease Research, Dhaka, Bangladesh; [4]Centro de Investigacao em Saude, Manhica, Maputo, Mozambique; [5]Kenya Medical Research Institute/CDC, Kisumu; [6]National Institute of Cholera and Enteric Diseases, Kolkata, India; [7]Center for Vaccine Development–Mali, Bamako; [8]The Aga Khan University, Karachi, Pakistan; [9]Department of Medicine, University of Virginia School of Medicine, Charlottesville; [10]US Centers for Disease Control and Prevention, Atlanta, Georgia; [11]Universidad de Chile, Santiago; [12]World Health Organization, Geneva, Switzerland; [13]Institut Pasteur, Paris, France; [14]University of Melbourne, Australia; [15]University of Gothenburg, Sweden; and [16]Department of Pediatrics, University of Virginia School of Medicine, Charlottesville

To understand the etiology of moderate-to-severe diarrhea among children in high mortality areas of sub-Saharan Africa and South Asia, we performed a comprehensive case/control study of children aged <5 years at 7 sites. Each site employed an identical case/control study design and each utilized a uniform comprehensive set of microbiological assays to identify the likely bacterial, viral and protozoal etiologies. The selected assays effected a balanced consideration of cost, robustness and performance, and all assays were performed at the study sites. Identification of bacterial pathogens employed streamlined conventional bacteriologic biochemical and serological algorithms. Diarrheagenic *Escherichia coli* were identified by application of a multiplex polymerase chain reaction assay for enterotoxigenic, enteroaggregative, and enteropathogenic *E. coli*. Rotavirus, adenovirus, *Entamoeba histolytica*, *Giardia enterica*, and *Cryptosporidium* species were detected by commercially available enzyme immunoassays on stool samples. Samples positive for adenovirus were further evaluated for adenovirus serotypes 40 and 41. We developed a novel multiplex assay to detect norovirus (types 1 and 2), astrovirus, and sapovirus. The portfolio of diagnostic assays used in the GEMS study can be broadly applied in developing countries seeking robust cost-effective methods for enteric pathogen detection.

Diarrheal diseases remain among the leading global causes of death for children <5 years of age. A major shortcoming of diarrheal disease studies conducted prior to The Global Enteric Multicenter Study (GEMS) has been the failure to perform a comprehensive ascertainment of major enteric pathogens, particularly at sites of greatest diarrheal burden. This deficit is understandable, considering that sites with high diarrheal mortality are typically those with the greatest challenges to performing the technically demanding portfolio of assays and protocols required to identify bacterial, viral, and protozoal pathogens. Thus, a goal of GEMS has been to assure accurate and consistent identification of relevant pathogens at all the GEMS study sites.

In order to accomplish the challenging but important task of identifying consistently the key pathogens at all GEMS sites, within the significant internal and external constraints, we established the following requirements for a comprehensive set of diagnostic tests:

1. Performance: The methods utilized were required to have satisfactory sensitivity and specificity. Although difficult to define, we aspired to achieve performance that equaled the standards necessary for effective clinical management in most settings, and

Correspondence: James P. Nataro, MD, PhD, MBA, Dept of Pediatrics, University of Virginia School of Medicine, 3601 West Clinics, Hospital Dr, Charlottesville, VA 22908 (jpn2r@virginia.edu).

Clinical Infectious Diseases 2012;55(S4):S294–302

DOI: 10.1093/cid/cis754

satisfactory to assure sufficiently accurate ascertainment of burden and the generation of reliable data.

2. Robustness: Although all of the sites introduced stringent quality assurance (QA) structures, the methods needed to be consistent across all the sites, requiring feasible training and oversight, as well as the opportunity for verification and validation using post hoc studies.
3. Cost-effectiveness: GEMS operated on a generous but limited budget. We were required to introduce assays that could be performed within reasonable financial constraints.
4. The Delphic perspective: We enlisted respected experts on each pathogen to ensure expert support in method selection, personnel training, and QA programs.

Herein we describe the clinical microbiology laboratory methods and protocols utilized in the GEMS study. Most of these assays were adapted from published methods that had independently been developed, validated and subjected to peer review.

Collection and Processing of Stool Samples

Fecal samples in the GEMS study were delivered to the laboratory in cold containers (see Kotloff et al in this supplement). Either at the point of collection or upon accession in the laboratory, a fecal aliquot was introduced into 2 tubes, one containing Cary-Blair medium [1] and one buffered glycerol saline (BGS) [2]. When no fecal specimen was available, a rectal swab was obtained; these rectal swabs were immediately inserted into tubes containing Cary-Blair and BGS media.

Upon arrival at the laboratory, the lab personnel inspected the sample for temperature and stool volume of at least 3 mL; an accession form was processed. The time between stool collection and inoculation of transport media needed to be not more than 6 hours, and the time between placing the specimen in transport media and accession was not more than 18 hours. Aliquots of stool samples were prepared and frozen for subsequent tests as described below.

Conventional Fecal Microbiology

The GEMS protocol included conventional bacterial culture, primarily so that pure growth of implicated pathogens could be independently validated by central laboratories and characterized further with regard to virulence, serologic, and antimicrobial resistance properties.

Bacteria selected for isolation and identification included gram-negative bacteria of proven or highly suspected pathogenicity and significance in developing world settings, as evidenced by the world's literature. The final list of agents sought was vetted through the investigators and the GEMS Microbiology Steering Committee. The pathogens sought included diarrheagenic (enterotoxigenic [ETEC], enteropathogenic [EPEC], and enteroaggregative [EAEC]) *Escherichia coli*, serovars of *Salmonella enterica*, *Shigella* spp, *Campylobacter* spp, *Vibrio* spp, *and Aeromonas* spp. The algorithm for bacteriologic characterization comprised a differential medium, a moderately selective medium, a highly selective medium, and at least 1 enrichment broth. All protocols were adapted from the *Manual of Clinical Microbiology*, Eighth Edition [3]. From the Cary-Blair tube, swabs were plated onto MacConkey (MAC), xylose lysine desoxycholate (XLD), thiosulfate citrate bile salts sucrose (TCBS), Aeromonas (Ryan) [4], Campy-BAP [5], and alkaline peptone water media; from the BGS the swab was plated onto MAC and XLD media. Plates were incubated at 37°C with the exception of media for *Campylobacter* spp (42°C) and *Aeromonas* spp (10°C–42°C). After incubation, suspicious colonies were selected and subjected to a series of simple biochemical tests that could be performed conveniently in resource-poor settings, minimizing expense, difficulty in procurement of reagents, and need for sophisticated training or equipment. The confirmatory tests utilized are described below.

Enterobacteriaceae

Colonies were inoculated into triple-sugar iron, motility indole ornithine (MIO), and lysine decarboxylase media, as well as citrate and urea biochemical typing media, and incubated at 35°C–37°C overnight. Isolates biochemically suspicious for *Salmonella enterica* [urea (−) oxidase (−)] were serotyped with polyvalent O and Vi following the manufacturer's instructions (Denka Seiken). All isolates biochemically identified as *Shigella* spp were serotyped with polyvalent group A, B, C, and D using manufacturer's protocols (Denka Seiken or Reagensia).

Vibrio spp Isolation and Identification

TCBS agar plates were examined for growth on day 2; large yellow and green colonies were subcultured to Trypticase soy agar (TSA) and incubated at 37°C overnight. When there was no growth of colonies resembling *Vibrio* spp after overnight incubation on the TCBS plates, subculture from TSA was tested for the production of oxidase; if oxidase negative, then no further for testing for *Vibrio* spp was done. If oxidase positive, the isolates were tested for salt tolerance with different concentration of NaCl supplemented in nutrient broth (0%, 6%, and 8%). If the colony was yellow on TCBS and there was growth in 0% and no growth at 8% NaCl-nutrient broth, then the putative *Vibrio* isolates were reincubated at 37°C for another 24 hours; at the same time the alkaline peptone water was subcultured to a new TCBS plate and incubated at 37°C. On day 3, each *Vibrio cholerae* was confirmed serologically using O1 and O139 antisera (Denka Seiken) and *V. cholerae* O1–positive cultures were typed as Inaba or Ogawa serotypes. If the colony was green on TCBS and there was growth in NaCl concentrations of 6% and 8%, and no growth in 0%, this was considered presumptive for *Vibrio parahaemolyticus*.

Aeromonas spp Isolation and Identification

On day 2, Ryan agar plates were examined for dark green colonies with darker green centers. Such colonies were subcultured onto TSA plates, tested for salt tolerance with different concentrations of NaCl (0%, 6% and 8%), and incubated aerobically for 24 hours. The next day, oxidase and catalase tests from the TSA plate were performed and tubes read for growth at various NaCl concentrations. Susceptibility to O/129 (2, 4-diamino-6, 7-diisopropyl pteridine) was also assessed [6]. Any isolate that was oxidase (+), catalase (+), grew in 0% NaCl but not in 6% or 8%, and was resistant to O/129, was considered to belong to the species *Aeromonas*.

Campylobacter spp Isolation and Identification

On day 3 the Campy blood agar plate was observed for growth appearing in one of the following ways: (1) nonhemolytic, gray, yellowish or pinkish tint; (2) flat, spreading, irregular edged colonies; (3) mucoid; (4) thin film; (5) spreading along the streak mark; or (6) round and convex. Oxidase and catalase tests were done and a sodium hippurate tube was inoculated. If isolates were oxidase (+) and catalase (+), smears were prepared for Gram staining. The smear was examined under the light microscope for small gram-negative rods that are slightly curved or "S" shaped. The sodium hippurate hydrolysis test was then performed for confirmation. Hippurate hydrolysis positive isolates were classified as *Campylobacter jejuni*; if hippurate hydrolysis was negative, strains were classified as *Campylobacter coli*.

E. coli Isolation and Identification

From 2-day growth on MAC plates, several lactose-fermenting bacterial colonies resembling *E. coli* were picked and tested using MIO medium. Up to 3 lactose-positive and indole-positive colonies were selected. When there were multiple distinct *E. coli*–like colony morphologies, each was selected. If there were <3 colonies of lactose-fermenting *E. coli*–like organisms, then all lactose-positive colonies were picked, and ≥1 lactose-negative colonies were picked to reach the total of 3 colonies per specimen. Indole-positive colonies were saved for further analysis. For indole-negative colonies, a second series of biochemical test, Indole/Methyl Red/Voges Proskauer/Citrate was used to identify *E. coli*. If any were positive for methyl red, and negative for Voges Proskauer and citrate, they were saved for further analysis. If 3 presumed *E. coli* were not found (ie, positive for indole or another suggestive biochemical reaction), the microbiologist returned to the original plate and picked up to 3 additional colonies for biochemical testing.

ETEC, EPEC, and EAEC pathotypes were identified using a multiplex polymerase chain reaction (PCR) previously published [7], but adapted for the purpose of GEMS. The targets sought via the PCR reaction included ETEC heat-labile enterotoxin and heat-stable enterotoxin (derived from STh) genes, the EPEC intimin (*eae* gene) outer membrane protein adhesin; the EPEC plasmid-encoded bundle-forming pilus (BFP); the EAEC plasmid-encoded gene *aatA*; and the EAEC chromosomally encoded *aaiC* locus. All of these loci are known virulence determinants of their respective pathogens [8, 9]. Strains positive for *eae* but not BFP were designated atypical EPEC. Strains positive for either ETEC enterotoxin were considered ETEC and strains positive for either EAEC factor were considered EAEC for the purposes of the GEMS analysis.

The 3 *E. coli*–like colonies selected from each stool were pooled into a common sample tube and template DNA was prepared from the pooled colonies. Template DNA was prepared by boiling the cultures grown on L-agar for 20 minutes, rapidly cooling on ice, followed by brief centrifugation at

Table 1. Primer Sequences and the Expected Amplicon Sizes for the Multiplex Polymerase Chain Reaction Employed in the Detection of Diarrheagenic *Escherichia coli*

Pathogen	Primer	Target Gene	Primer Sequence (5′-3′)	Amplicon (bp)
ETEC	LT-F	*elt*	CACACGGAGCTCCTCAGTC	508
	LT-R		CCCCCAGCCTAGCTTAGTTT	
	ST-F	*est*	GCTAAACCAGTAG/AGGTCTTCAAAA	147
	ST-R		CCCGGTACAG/AGCAGGATTACAACA	
EPEC	BFPA-F	*bfpA*	GGAAGTCAAATTCATGGGGG	367
	BFPA-R		GGAATCAGACGCAGACTGGT	
	EAE-F	*eae*	CCCGAATTCGGCACAAGCATAAGC	881
	EAE-R		CCCGGATCCGTCTCGCCAGTATTCG	
EAEC	CVD432F	*aatA*	CTGGCGAAAGACTGTATCAT	630
	CVD432R		CAATGTATAGAAATCCGCTGTT	
	AAIC F	*aaiC*	ATTGTCCTCAGGCATTTCAC	215
	AAIC R		ACGACACCCCTGATAAACAA	

Abbreviations: EAEC, enteroaggregative *Escherichia coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*.

2500*g* for 10 minutes. This supernatant was used in the PCR assays. Primer nucleotide sequences and the predicted lengths of the resulting amplicons are listed in Table 1.

For the PCR reaction, 3 μL of template DNA was added to the PCR mix containing 2.5 μL of 10× PCR buffer with 2 mM $MgCl_2$ (New England Biolabs), 2.0 μL of 10 mM deoxynucleotide triphosphates (dNTPs) (Fermentas), 0.4 μL of 20 pmol/μL of each primer, 0.25 μL Taq DNA polymerase (5 U/μL, New England Biolabs), and 7.37 μL RNase-free water to a final volume of 20 μL. PCR was performed under the following conditions: preheating at 96°C for 4 minutes, denaturation at 95°C for 20 seconds, annealing at 57°C for 20 seconds, elongation at 72°C for 1 minute. PCR was performed for 35 cycles with final extension at 72°C for 7 minutes in an Eppendorf Mastercycler Gradient thermal cycler. The same model thermal cycler was employed at all sites. The amplification products were separated through a 2% agarose gel and visualized by ultraviolet light transillumination after ethidium bromide staining. The 1-kb plusA 100-bp DNA ladder (New England Biolabs) was used as a molecular size marker in gel. Appearance of the PCR amplicons on agarose gel electrophoresis is shown in Figure 1. Control strains employed in every PCR reaction were ETEC H10407, EAEC 042, and for EPEC strains CVD 28 (*eae*-positive) and HB101(pMAR7) (*bfpA*-positive).

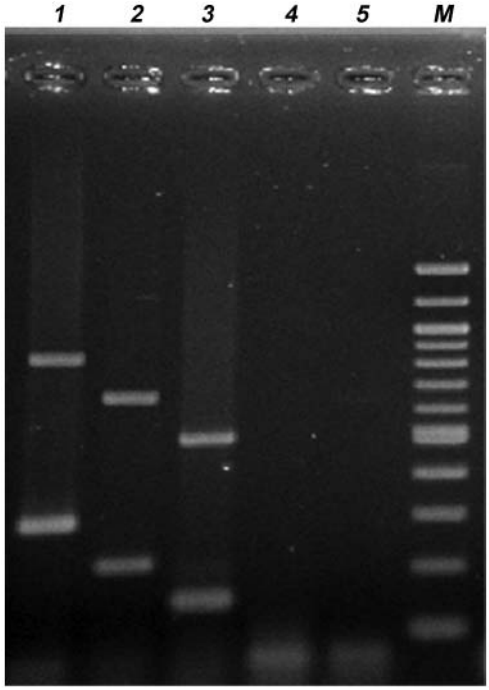

Figure 1. Appearance of diarrheagenic *Escherichia coli* amplicons separated by agarose gel electrophoresis. Lane 1, enteropathogenic *E. coli*; lane 2, enteroaggregative *E. coli*; 3, enterotoxigenic *E. coli*; lanes 4 and 5, negative control strains; lane 6, 100 bp DNA ladder (New England Biolabs).

Characterization of *eae*-Positive, *bfpA*-Negative Strains

As part of a nested study, all *E. coli* specimens that were negative in the original multiplex PCR for *elt*, *est*, *bfpA*, *eae*, *aatA*, and *aaiC* were investigated at the University of Melbourne, Australia, for *eae* by using a high-throughput real-time PCR assay. Specimens, consisting of 3 individual isolates, were sent to Melbourne from Baltimore on MAC agar in 96-well flat-bottomed microtiter trays. Upon arrival, the cultures were replica-plated onto MAC agar and grown overnight at 37°C. To generate template DNA for use in the real-time PCR, a sterile pipette tip was used to transfer a portion of a culture sample from the MAC replica plate into a single well of a 96-well PCR tray (Bio-Rad) containing 100 μL DNase-free water. This procedure was repeated for the remaining 2 samples of the specimen, so that each well contained 1 specimen comprising 3 separate isolates. The plate was sealed with Microseal "A" adhesive (Bio-Rad). To lyse the bacterial cells, the samples were heated to 99°C for 10 minutes in a C1000 PCR machine (Bio-Rad) followed by cooling at 12°C. Before use the plate was centrifuged for 1 minute at 3000*g* and the supernatant was used as the template DNA in the real-time PCR assay.

For the real-time PCR, 8 μL of a master mix was added to individual wells of a 96-well PCR tray (Bio-Rad) followed by 2 μL of template DNA. The real-time master mix, for one reaction, comprised 5.0 μL of 2 × SSoFast EvaGreen Supermix (Bio-Rad), 1.4 μL DNase-free water, and 0.8 μL of 5 μM of each primer. The plate was sealed with Microseal "B" adhesive (Bio-Rad) and centrifuged for 30 seconds at 3000*g*. Real-time PCR was performed using a CFX96 real-time PCR machine (Bio-Rad) using the following protocol: 95°C for 2 minutes, followed by 35 cycles of 95°C for 1 second, and 60°C for 5 seconds. The duration of one complete reaction was 24 minutes and upon completion the results were analyzed using

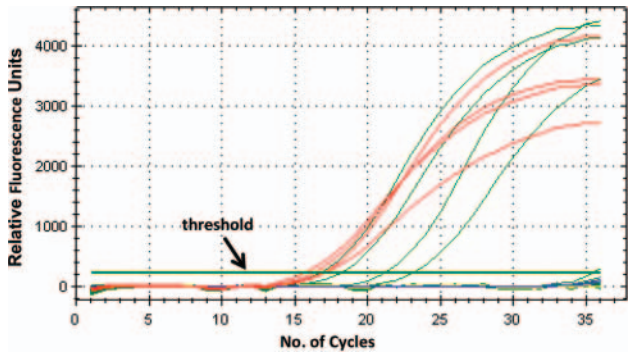

Figure 2. An example of the graphical results of real-time polymerase chain reaction performed on 4 *eae*-positive specimens (red), 4 unknown specimens (green), and negative controls (yellow and blue). A threshold for detection of DNA-based fluorescence is set slightly above background fluorescence levels.

the CFX Manager Software (Bio-Rad). Binding of the SSoFast EvaGreen dye to double-stranded DNA PCR products causes the dye to fluoresce. The cycle threshold is the number of cycles at which the fluorescence exceeds the background level (Figure 2). In our study, specimens with a cycle threshold of ≤30 were analyzed further. Control strains employed in every PCR included EPEC strains E2348/69, E128010, W1056, and TR952, which carry intimin alpha, beta, gamma, and epsilon, respectively (positive controls); and ETEC strain H10407 and *E. coli* K-12 strain MC4100 (negative controls). Three "no DNA template" controls were also included. Each individual isolate within an *eae*-positive specimen was analyzed by using a multiplex PCR to confirm the presence of *eae*, and also to test for the presence of genetic markers of typical EPEC (*bfpA*), Shiga toxin–producing *E. coli*, and/or enterohemorrhagic *E. coli* (EHEC) (*stx1*, *stx2*, *ehxA*). Template DNA for use in this PCR was prepared by resuspending a loopful of the individual culture samples from the MAC replica plate in 500 μL of DNase-free water and then boiling the suspension for 10 minutes. The boiled bacterial lysate was rapidly cooled on ice for 5 minutes followed by centrifugation for 5 minutes at 16 000*g*. The supernatant containing the DNA was transferred to a fresh microfuge tube and placed on ice or at 4°C until used in the PCR.

For this PCR a GoTaq Green Master Mix (Promega), which contained Taq DNA polymerase, dNTPs, $MgCl_2$, reaction buffers, and loading dye, was used. The PCR was performed in a C1000 PCR machine (Bio-Rad) using the following protocol: 95°C for 5 minutes, followed by 35 cycles of 95°C for 20 seconds, 55°C for 45 seconds, and 72°C for 30 seconds, followed by 1 cycle of 72°C for 7 minutes. The amplification products were separated through a 2% Tris-acetate-EDTA

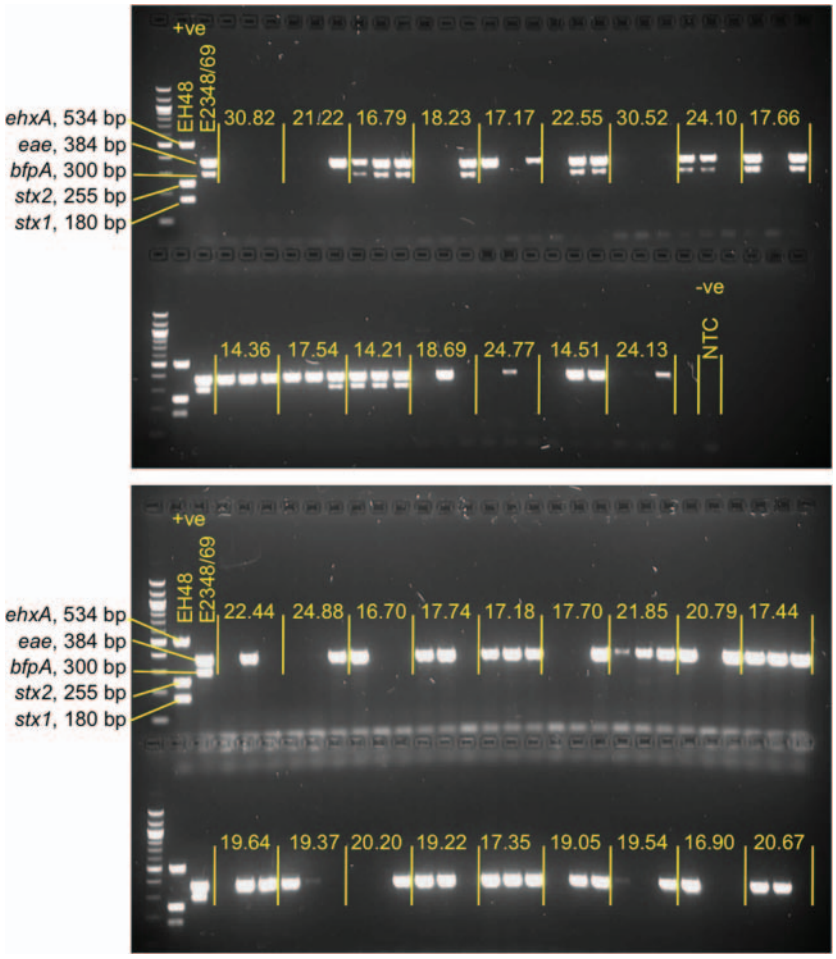

Figure 3. Gels showing the results of a multiplex polymerase chain reaction (PCR) assay for enteropathogenic *Escherichia coli* (EPEC), Shiga toxin–producing *E. coli*, and enterohemorrhagic *E. coli* (EHEC). Individual isolates from 34 specimens were subjected to a multiplex PCR as described in the text. Each specimen, separated by yellow vertical lines, consists of 3 individual isolates. The yellow values indicate the cycle threshold obtained for each specimen in the real-time PCR used in the initial screening for *eae*. The amplicons produced by the positive controls, EPEC E2348/69 (*eae* and *bfpA*) and EHEC EH48 (*stx1*, *stx2*, and *ehxA*) are also shown. 100 bp DNA ladder was used as a molecular size marker. Abbreviation: NTC, no template control.

Table 2. Primer Sequences and Expected Amplicon Size for Real-time Polymerase Chain Reaction

Primer	Sequence (5′-3′)	Target Gene	Amplicon (bp)
eae83-F	CAGGCTTCGTCACAGTTG	*eae*	83
eae83-R	CCGTCAAAGTTATTACCACTCTG		

agarose gel and visualized by ultraviolet light transillumination. A 100-bp DNA ladder (New England Biolabs) was used as a molecular size marker. Examples of the results of this PCR are shown in Figure 3. Control strains included in every PCR reaction were EPEC strain E2348/69 for *eae* and *bfpA* and EHEC strain EH48 for *stx1*, *stx2*, and *ehxA*. Primers that were used in the *eae* real-time PCR are listed in Table 2; reaction conditions are listed in Table 3; primer nucleotide sequences and the predicted lengths of the resulting amplicons are listed in Table 4.

Virus Immunoassays

Enzyme immunoassays are rapid, robust, sensitive, and specific diagnostic assays for some viral pathogens. We used well-validated commercial immunoassays for rotavirus and adenovirus according to established protocols.

Rotavirus

Rotavirus VP6 antigen was detected in stools by the ProSpecT ELISA Rotavirus kit following the manufacturer's instructions (Oxoid).

Adenovirus

General adenovirus hexon protein was detected using ProSpecT Adenovirus Microplate assays according to the manufacturer's instructions (Oxoid). This test utilizes a genus-specific monoclonal antibody to detect epitopes common to all human adenovirus serotypes.

Samples for adenovirus by the ProSpecT assay were further tested for the presence of enteric adenovirus serotypes 40/41 using Premier Adenoclone kit (Meridian Bioscience) following the manufacturer's instructions.

Table 3. Components of the Multiplex Polymerase Chain Reaction

1 × polymerase chain reaction	
2 × GoTaq Green Master Mix	15.0 µL
20 µM bfpA-F	1.0 µL
20 µM bfpA-R	1.0 µL
20 µM ehxA-F	1.0 µL
20 µM ehxA-R	1.0 µL
20 µM eae-F	1.0 µL
20 µM eae-R	1.0 µL
20 µM stx1-F	0.5 µL
20 µM stx1-R	0.5 µL
20 µM stx2-F	0.5 µL
20 µM stx2-R	0.5 µL
DNA template	2.0 µL
Total volume	25.0 µL

Table 4. Primer Sequences and the Expected Amplicon Sizes for the Multiplex Polymerase Chain Reaction

Primer	Sequence (5′-3′)	Target Gene	Amplicon (bp)
eae-F	GACCCGGCACAAGCATAAGC	*eae*	384
eae-R	CCACCTGCAGCAACAAGAGG		
ehxA-F	GCATCATCAAGCGTACGTTCC	*ehxA*	534
ehxA-R	AATGAGCCAAGCTGGTTAAGCT		
stx1-F	ATAAATCGCCATTCGTTGACTAC	*stx1*	180
stx1-R	AGAACGCCCACTGAGATCATC		
stx2-F	GGCACTGTCTGAAACTGCTCC	*stx2*	255
stx2-R	TCGCCAGTTATCTGACATTCTG		
bfpA-F	GGAAGTCAAATTCATGGGGG	*bfpA*	300
bfpA-R	GGAATCAGACGCAGACTGGT		

Multiplex PCR for Detection of RNA Viruses

Stool specimens were diluted to 10% (w/v or v/v) suspensions in Vertrel XF (Miller Stephenson) and centrifuged at 1000*g* for 10 minutes. The supernatant was collected and stored at 4° C prior to RNA extraction.

Viral RNA was extracted from stool supernatant using Nuclisens (bioMérieux) as per the manufacturer's instructions. In brief, 900 µL of lysis buffer was added to 200 µL of supernatant, vortexed and incubated for 10 minutes, then 50 µL of silica suspension was added, vortexed and centrifuged at 10 000*g* for 30 seconds. Washing was done by adding 1 mL of wash buffer twice followed by washing with 1 mL of 70% ethanol twice. Finally 1 mL of acetone was added to the pellet. At the end of each washing step, tubes were vortexed and centrifuged at room temperature for 30 seconds at 10 000*g*; supernatant was carefully discarded without disturbing silica pellet. The silica pellet was dried at 56°C for 10 minutes and the pellet was reconstituted by adding 50 µL of elution buffer. Samples were vortexed and incubated at 56°C for 5 minutes, the incubation step was repeated, and the specimen was centrifuged for 2 minutes at 10 000*g*. RNA containing supernatant was collected containing RNA and stored at −70°C until use.

RNA was reverse transcribed in a total volume of 15 µL containing 1× First strand buffer (Invitrogen), 0.5 mM dNTPs (Roche), 0.5 mM dithiothreitol (Invitrogen), 0.5 µg of random primers (TaKaRa), 20 units of RNase Inhibitor (Roche), and

Table 5. Primer Sequences and the Expected Amplicon Sizes for the Multiplex Polymerase Chain Reaction Used in the Detection of RNA Viruses

Pathogen	Primer	Primer Sequence (5′-3′)	Amplicon (bp)
Norovirus GI	G1SKR	CCAACCCARCCATTRTACA	330
	G1SKF	CTGCCCGAATTYGTAAATGA	
Norovirus GII	G2SKR	CCRCCNGCATRHCCRTTRTACAT	387
	COG2F	CARGARBCNATGTTYAGRTGGATGAG	
Sapovirus	SLV5749	CGGRCYTCAAAVSTACCBCCCCA	434
	SLV5317	CTCGCCACCTACRAWGCBTGGTT	
Astrovirus	82b	GTGAGCCACCAGCCATCCCT	719
	PreCAP1	GGACTGCAAAGCAGCTTCGTG	
	Cog2R	TCGACGCCATCTTCATTCACA	

150 units of Superscript II Reverse Transcriptase (RT; Invitrogen). The mixture was incubated at 42°C for 1 hour and then heated at 99°C for 5 minutes.

A multiplex PCR reaction was designed to amplify norovirus, astrovirus, and sapovirus complementary DNA (cDNA) present in the reverse transcription reactions described above. The method was adapted from a published protocol [10]. After cDNA synthesis, multiplex PCR was performed using specific primers (Table 5). PCR master mix contained 0.5 μM concentration of specific primers, 0.2 mM dNTPs (Roche), 1× AmpliTaq buffer I, and 1.25 U of AmpliTaq DNA polymerase (Applied Biosystems) for a 25-μL reaction. Master Mix was distributed to 0.2-mL PCR tubes, and 5 μL of template cDNA was added. The assay was confirmed using positive and negative controls cDNA from confirmed prior reactions. PCR reactions were conducted in a Eppendorf Mastercycler Gradient thermal cycler starting with a denaturing step of 3 minutes at 94°C, followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C, followed by an extension of 72°C for 7 minutes. After the thermocycling step, all PCR products were electrophoresed on a 2.0% agarose gel and sized with a 100-bp ladder (Promega) (Figure 4).

Detection of Protozoal Pathogens

Giardia enterica [11, 12], *Entamoeba histolytica* [13], and *Cryptosporidium* spp [12] were detected using immunoassays available commercially from TechLab, Inc and according to manufacturer's protocols. Studies have demonstrated excellent performance of these assays, superior to microscopic detection [14–17].

Quality Control Methods

Initial Training

An investigators' meeting was held at the start of the study at the Center for Vaccine Development (CVD) in Baltimore, to review the procedures to be used. All the laboratory heads from the field sites and some technicians attended the meeting. CVD Quality Control (QC)/QA staff reviewed the execution of each standard operating procedure (SOP) during site visits and provided retraining if necessary.

Standard Operating Procedures

In order to streamline processes at each site, SOPs were generated to ensure that all procedures were executed in consistent fashion at each site. SOPs clearly defined the purpose, the required materials and equipment, safety guidelines,

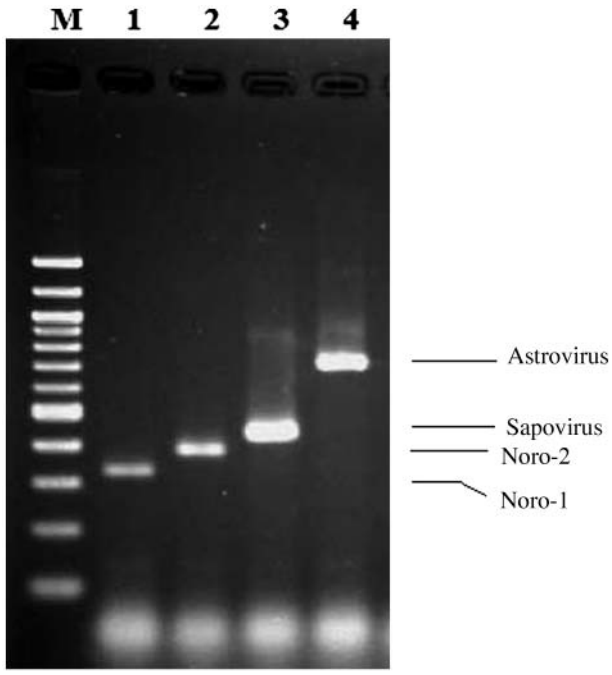

Figure 4. Appearance of enteric viral amplicons separated by agarose gel electrophoresis. Lane M, 100 bp DNA ladder (New England Biolabs); lane 1, Norovirus GI (330 bp); lane 2, Norovirus GII positive (387 bp); lane 3, sapovirus (434 bp); lane 4, astrovirus (719 bp).

responsibility, procedures, and documentation and provided related documents.

By introducing controlled forms for each SOP, a "quality checklist" was created that would ensure that each step in the SOP was carried out as directed, and that the materials used were as stipulated in the SOP and prior to their expiration dates. These forms were also reviewed by the laboratory supervisor or designee to ensure adherence to the SOP and that quality deliverables were generated. Forms also included, for some tests, negative, positive, and cutoff values. Samples that were not valid, or test runs in which control values were not valid, were repeated. The rate of sample reworking was tracked as a quality metric.

Quality Assurance

Quality incidents and deviations from the SOP were reported and documented on designated forms and reviewed by the supervisor on site and by QC/QA personnel during regular site visits. Corrective and preventive actions were executed on site, by the laboratory supervisor or designee. Very few or no quality incidents or deviations occurred for each protocol. All forms were reviewed by the QC/QA CVD staff during routine site visits.

All case report forms were also reviewed by the data coordinating center (DCC) for completeness. Missing data and/or missing forms were communicated to the sites via email. Other information, such as ranges of time, was also calculated by the DCC.

Biannual proficiency testing was conducted at each site. Sites were expected to score 80% on identification of "unknown" samples sent from Baltimore. All sites attained this score. Any incorrect results were investigated and any errors corrected and retraining provided if necessary on site by the laboratory managers.

Post Hoc Studies and Validation Studies

The GEMS study has generated a cornucopia of bacterial strains, fecal nucleic acid, and frozen stool strains that will yield priceless information regarding the agents associated with diarrhea in infants and young children in developing countries and their genomic and serologic diversity. The analyses proposed in the GEMS protocol include typing of the major ETEC adhesins, the colonization factors. In addition, *Shigella dysenteriae* isolates were tested to detect *S. dysenteriae* 1 (the Shiga bacillus), all *Shigella sonnei* were serologically confirmed, and all *Shigella flexneri* isolates were typed and subtyped. These 2 analyses will profile the antigenic diversity of these 2 important pathogens and inform future vaccine development priorities and will be reported elsewhere.

The availability of GEMS clinical samples also provides the opportunity for diagnostic method development and validation. For example, a rigorous comparison of the multiplex RT-PCR assay with real-time PCR for detection of norovirus has been completed and will be described elsewhere. In addition, the sample archive provides the platform for the development of new, high-throughput and highly multiplexed diagnostic technologies, comparing their performance with gold standard methodologies.

DISCUSSION

The GEMS study employed a portfolio of diagnostic tests that balanced practicality and economy, as well as good sensitivity and specificity. A number of important issues warrant elaboration.

We decided to employ conventional bacteriologic methods for isolation of putative bacterial pathogens, followed by molecular and/or phenotypic characterization. The derivation of pure bacterial stocks permitted not only downstream characterization of genetic and surface markers of relevance to epidemiology and vaccine development, but also allowed us to revisit the diagnostic performance of the selected assays on archived strain collections. *Escherichia coli* colonies, for example, were isolated, archived, and tested for the presence of virulence-related genes that define diarrheagenic pathotypes; *Shigella* strains were serotyped at reference laboratories in order to inform future vaccine development strategies. As noted, validation of both EPEC and ETEC primers sets was performed on the *E. coli* archive using high-throughput PCR analysis. For EAEC, which was not associated with diarrhea overall, the availability of archived bacterial cultures permitted extensive genomic characterization of isolates, thereby identifying potentially pathogenic genotypes [18].

Agarose gel–based detection of PCR amplicons was the preferred in the GEMS diagnostic set for the following reasons. At the time the GEMS protocol was developed, there was little expertise in real-time PCR at any of the sites in the GEMS network, and the added complexity of real-time was beyond what the training programs could realistically accomplish. Additional advantages of the gel-based method include substantially lower cost, the availability of gel images that could be shared across sites for validation and quality control purposes, and greater availability of supplies at the sites.

We decided to employ immunoassays for detection of protozoal pathogens for many of the same reasons. Direct microscopic detection of protozoal pathogens requires significant expertise and is not readily amenable to downstream validation. Immunoassays, also employed for detection of some viral agents, followed a simple, highly standardized, and centrally validated method that was easily deployed at the study sites. An additional advantage to enzyme immunoassay methods was the availability of product support from the kit manufacturers.

GEMS investigators applied multiple criteria by which to select agents for detection. These criteria included published citation as a significant agent of childhood diarrhea at multiple sites in the developing world, and practical detection methodology. Toxigenic *Bacteroides fragilis*, for example, could have been included in the portfolio but would have required either anaerobic bacteriology or use of tests that could not be validated post hoc on pure cultures. The availability of the GEMS specimen archive permits post hoc detection of additional agents using molecular and other technologies, and these efforts are under way.

All primers employed in PCR reactions were selected from published studies, thereby conferring both validation by an independent laboratory and peer review, and were also validated in the laboratories of the GEMS investigators in Baltimore. Post hoc validation was nevertheless carried out employing nested studies of individual block PCR reactions and/or the use of alternative primer sets.

The GEMS study offers a quantum leap in our understanding of the burden and etiology of diarrhea afflicting infants and young children in developing countries. The GEMS etiology data and specimen collections will be grist for further advances far into the future.

Notes

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Cary SG, Blair EB. New transport medium for shipment of clinical specimens. I. Fecal specimens. J Bacteriol **1964**; 88:96–8.
2. Wells JG, Morris GK. Evaluation of transport methods for isolating *Shigella* spp. J Clin Microbiol **1981**; 13:789–90.
3. Murray PR, Baron EJ, Jorgensen JH, Pfaller MA, Yolken RH. Manual of clinical microbiology. Washington, DC: American Society for Microbiology, **2003**.
4. Bernagozzi M, Bianucci F, Scerre E, Sacchetti R. Assessment of some selective media for the recovery of *Aeromonas hydrophila* from surface waters. Zentralbl Hyg Umweltmed **1994**; 195:121–34.
5. Blaser MJ, Berkowitz ID, Laforce FM, Cravens J, Reller LB, Wang WL. *Campylobacter* enteritis: clinical and epidemiologic features. Ann Intern Med **1979**; 91:179–85.
6. Khaitovich AB, Ved'mina EA. [Vibrio and aeromonad sensitivity to the vibriostatic O 129.] Antibiot Med Biotekhnol **1987**; 32:446–9.
7. Nguyen TV, Le VP, Le HC, Gia KN, Weintraub A. Detection and characterization of diarrheagenic *Escherichia coli* from young children in Hanoi, Vietnam. J Clin Microbiol **2005**; 43:755–60.
8. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2004**; 2:123–40.
9. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **1998**; 11:142–201.
10. Yan H, Yagyu F, Okitsu S, Nishio O, Ushijima H. Detection of norovirus (GI, GII), sapovirus and astrovirus in fecal samples using reverse transcription single-round multiplex PCR. J Virol Methods **2003**; 114:37–44.
11. Boone JH, Wilkins TD, Nash TE, et al. TechLab and Alexon *Giardia* enzyme-linked immunosorbent assay kits detect cyst wall protein 1. J Clin Microbiol **1999**; 37:611–4.
12. Youn S, Kabir M, Haque R, Petri WA Jr. Evaluation of a screening test for detection of *Giardia* and *Cryptosporidium* parasites. J Clin Microbiol **2009**; 47:451–2.
13. Blessmann J, Ali IK, Nu PA, et al. Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. J Clin Microbiol **2003**; 41:4745–50.
14. Bialek R, Binder N, Dietz K, Joachim A, Knobloch J, Zelck UE. Comparison of fluorescence, antigen and PCR assays to detect *Cryptosporidium parvum* in fecal specimens. Diagn Microbiol Infect Dis **2002**; 43:283–8.
15. Gaafar MR. Evaluation of enzyme immunoassay techniques for diagnosis of the most common intestinal protozoa in fecal samples. Int J Infect Dis **2011**; 15:e541–4.
16. Haque R, Ali IK, Akther S, Petri WA Jr. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J Clin Microbiol **1998**; 36:449–52.
17. Srijan A, Wongstitwilairoong B, Pitarangsi C, et al. Re-evaluation of commercially available enzyme-linked immunosorbent assay for the detection of *Giardia lamblia* and *Cryptosporidium* spp from stool specimens. Southeast Asian J Trop Med Public Health **2005**; 36(suppl 4):26–9.
18. Boisen N, Scheutz F, Rasko DA, et al. Genomic characterization of enteroaggregative *Escherichia coli* from children in Mali. J Infect Dis **2012**; 205:431–44.

SUPPLEMENT ARTICLE

Factors That Explain Excretion of Enteric Pathogens by Persons Without Diarrhea

Myron M. Levine[1] and Roy M. Robins-Browne[2]

[1]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore; and [2]Department of Microbiology and Immunology, The University of Melbourne, and Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

Excretion of enteropathogens by subjects without diarrhea influences our appreciation of the role of these pathogens as etiologic agents. Characteristics of the pathogens and host and environmental factors help explain asymptomatic excretion of diarrheal pathogens by persons without diarrhea. After causing acute diarrhea followed by clinical recovery, some enteropathogens are excreted asymptomatically for many weeks. Thus, in a prevalence survey of persons without diarrhea, some may be excreting pathogens from diarrheal episodes experienced many weeks earlier. Volunteer challenges with *Vibrio cholerae* O1, enterotoxigenic *Escherichia coli* (ETEC), enteropathogenic *E. coli*, *Campylobacter jejuni*, and *Giardia lamblia* document heterogeneity among enteropathogen strains, with some inexplicably not eliciting diarrhea. The immune host may not manifest diarrhea following ingestion of a pathogen but may nevertheless asymptomatically excrete. Some human genotypes render them less susceptible to symptomatic or severe diarrheal infection with certain pathogens such as *Vibrio cholerae* O1 and norovirus. Pathogens in stools of individuals without diarrhea may reflect recent ingestion of inocula too small to cause disease in otherwise susceptible hosts or of animal pathogens (eg, bovine or porcine ETEC) that do not cause human illness.

Clinical studies of 2 different designs, case/control and prospective longitudinal follow-up of a cohort, have historically played important roles in (1) identifying putative new diarrheal pathogens; (2) assessing the degree of pathogenicity of new or established enteropathogens; and (3) estimating the relative burden of different enteric pathogens. In case/control studies, clinical specimens from patients with diarrhea (cases) and properly matched (eg, by age and sex) control subjects without diarrhea are examined to detect the pathogens of interest. Odds ratios (ORs) are calculated to quantify the degree of association of the pathogen of interest with diarrhea. This involves comparing the odds of finding the pathogen in cases with the odds of finding the pathogen in controls; the higher the OR, the stronger the association. As described in the paper by Blackwelder et al in this supplement, the OR is also one key factor in the equation used to calculate the attributable fraction (AF) of a pathogen in a case/control study, thereby elucidating the relative contributions of different enteropathogens to the burden of diarrheal illness. Further statistical methods are employed to adjust for the presence of other enteric pathogens in the cases and controls [1]. When applying these statistical methods, it is evident that the prevalence rate of an enteropathogen in controls influences the estimates. The higher the rate of detection of the enteric pathogen(s) of interest in controls, the weaker the OR association (for pathogenicity) or the smaller the AF for that pathogen as a cause of diarrheal disease at the population level.

Similarly, when cohorts of children or adults are followed prospectively for the occurrence of diarrheal illness, the rate of detection of various pathogens of interest when a subject develops diarrhea is typically compared to serial "routine" specimens from that

Correspondence: Myron M. Levine, MD, DTPH, Center for Vaccine Development, University of Maryland School of Medicine, 685 W Baltimore St, Baltimore, MD 21201 (mlevine@medicine.umaryland.edu).

Clinical Infectious Diseases 2012;55(S4):S303–11

DOI: 10.1093/cid/cis789

subject that were collected systematically when he/she did not have diarrhea [2–4]. A "hybrid" approach is to nest a case/control strategy within the cohort study. Thus, a subject within the cohort who develops diarrhea is matched (usually by age and sex) to another subject within the cohort who at the time is free of diarrhea [5, 6]. In these cohort study strategies, the rate of detection of pathogens in stool specimens from the diarrhea cases is compared, respectively, to the rate of pathogen detection in the routine stool specimens from that person or in specimens from the matched control in the nested case/control approach. In these designs, as well, the rate of isolation of pathogens from the controls (or from the period when the subject is free of diarrheal illness) influences the conclusions that can be drawn about the pathogenicity of specific pathogens or their relative importance compared to other pathogens (as calculated using AF).

Finally, for clinicians who must make judgments about the need for specific therapeutic interventions based on the isolation of a specific diarrheal pathogen from a case of diarrhea, knowledge (from epidemiologic studies) of the relative frequency with which that enteric pathogen is found in healthy subjects without diarrhea provides information that may be helpful in decision making in the clinical situation.

Because the excretion of enteric pathogens in subjects without diarrhea influences our appreciation of the role of those pathogens as causes of diarrhea, it is imperative to consider the reasons why one finds diarrheal pathogens in healthy persons not suffering from diarrhea. Herein we review the characteristics of the pathogens, host factors, and environmental factors that provide explanations for the asymptomatic excretion of known diarrheal pathogens.

CHARACTERISTICS OF THE PATHOGEN

Unusually Long Duration of Excretion After Causing Diarrheal Illness

When subjects recover clinically following diarrheal illness caused by certain pathogens, the pathogens continue to be excreted asymptomatically for an extended period. Thus, when subjects without diarrhea are selected to serve as nondiarrheal controls, some may still be excreting a pathogen consequent to an episode of clinical diarrhea that may have occurred many weeks earlier. Enteric pathogens associated with extended excretion following an episode of acute diarrhea include nontyphoidal *Salmonella* [7, 8], *Campylobacter jejuni* [9–12], norovirus GI and GII [13–16], and, uncommonly, *Shigella* [17].

Heterogeneity of Pathogenicity Among Strains of the Pathogen

Experimental challenge studies in healthy adult volunteers who were fed various strains of known or putative enteric pathogens revealed that some strains caused diarrhea more readily than others at the same challenge inoculum, with some strains failing to cause diarrhea at all. Moreover, among the strains that did elicit diarrhea, the severity and range of symptoms sometimes varied widely. These observations were made with experimental challenge studies involving strains of *Vibrio cholerae* O1, enteropathogenic *Escherichia coli* (EPEC), enterotoxigenic *E. coli* (ETEC), *Campylobacter jejuni*, and *Giardia lamblia*. Thus, with many enteropathogens there appears to be heterogeneity among the strains that are circulating in human populations, with some strains being more prone to cause clinical disease than others. When many of these observations were initially made, the virulence attributes and other characteristics that differentiated the "diarrheagenic" strains from the other strains were not readily appreciated; in some instances the explanations are still not available.

In the early years following the identification of ETEC as pathogens, 3 broad categories came to be recognized, with some producing both heat-labile enterotoxin (LT) and heat-stable enterotoxin (ST), while others elaborated only ST or only LT [18]. Early clinical challenge studies showed that LT/ST strains [19, 20] and ST-only strains [21] reliably elicited watery diarrhea in volunteers. In contrast, LT-only strains were inconsistent in inducing diarrheal illness. LT-only strain E2528-C1, which was epidemiologically incriminated as responsible for an outbreak of acute diarrhea on a cruise ship [22], induced diarrheal illness after a relatively short incubation period when fed to volunteers [20]. In contrast, *E. coli* strain H10407P, which was derived from strain H10407 consequent to the loss of a plasmid encoding fimbrial colonization factor antigen I (CFA/I) and ST, did not cause diarrhea in volunteers even though the strain elaborated LT [23, 24], and the parent LT/ST, CFA/I-positive strain induced copious watery diarrhea [23–25]. These clinical trials provided early indications that fimbrial colonization factors play an important role in the pathogenesis of ETEC diarrhea in humans, as they do in ETEC pathogens of piglets and calves.

As shown in Table 1, similar experiences were observed when several different strains of *V. cholerae* O1 El Tor [26], EPEC [27], *C. jejuni* [28], and *G. lamblia* [29, 30] were fed to volunteers, even though all the strains were all isolated from patients with diarrheal illness. Thus, *V. cholerae* O1 El Tor strains N16961 and E7946, EPEC strains E2348/69 and E851/71, *C. jejuni* strain 81–176, and *G. lamblia* strain Gsm caused higher attack rates and more severe diarrhea, whereas *V. cholerae* O1 El Tor strain N16117, EPEC strain E74/68, *C. jejuni* strain A3249, and *G. lamblia* strain Isr either did not cause diarrhea or elicited lower attack rates or markedly milder clinical illness. Thus, in case/control studies of diarrhea in developing countries, it is possible that a proportion of controls with asymptomatic infection are carrying nonpathogenic or less pathogenic strains such as *V. cholerae* O1 N16117, EPEC

Table 1. Experimental Challenge Studies in Volunteers Documenting Variability in the Pathogenicity of Circulating Strains of Bacterial and Protozoal Pathogens Isolated From Patients With Diarrhea

Enteric Pathogen	Challenge Strain	Dose (CFU for Bacteria; No. of Trophozoites for Protozoa)	Diarrhea Attack Rate (%)	Positive Stool Culture or Pathogen Detection (%)	Ref.
Vibrio cholerae O1 El Tor	Inaba N16961	10^5	3/5 (60)	4/5 (80)	[26, 85]
	Inaba N16961	10^6	9/10 (90)	10/10 (100)	[26, 85]
	Ogawa E7946	10^6	6/6 (100)	6/6 (100)	[86]
	Ogawa N15870	10^5	3/5 (60)	4/5 (80)	[26, 85]
	Ogawa N15870	10^6	8/8 (100)	8/8 (100)	[26, 85]
	Ogawa N16117	10^5	0/4 (0)	2/4 (50)	[26, 85]
	Ogawa N16117	10^6	0/5 (0)	1/5 (20)	[26, 85]
EPEC	E851/71 (O142:H6)	10^{10}	5/5 (100)	5/5 (100)	[27]
	E2348/69 (O127:H6)	10^{10}	3/5 (60)	5/5 (100)	[27]
	E2348/69 (O127:H6)	10^{10}	11/11 (100)	11/11 (100)	[87]
	E74/68 (O128:H2)	10^{10}	0/5 (0)	5/5 (100)	[27]
ETEC	B2C (O6:H16)	10^8	2/5 (40)[a]	5/5 (100)	[19]
	B2C (O6:H16)	10^{10}	3/5 (60)[b]	5/5 (100)	[19]
	B7A (O148:H28)	10^8	1/5 (20)[a]	4/5 (80)	[19]
	B7A (O148:H28)	10^{10}	4/5 (80)[b]	5/5 (100)	[19]
	B7A (O148:H28)	10^6	3/6 (50)	6/6 (100)	[20]
	B7A (O148:H28)	10^8	7/11 (64)	11/11 (100)	[20]
	263 (pig strain)	10^8	0/5 (0)	5/5 (100)	[19]
	263 (pig strain)	10^{10}	0/5 (0)	3/4 (75)	[19]
Campylobacter jejuni	81–176	10^6	3/7 (43)	7/7 (100)	[28]
	81–176	10^8	6/10 (60)	10/10 (100)	[28]
	A3249	10^6	2/19 (11)	15/19 (79)	[28]
	A3249	10^8	0/5 (0)	5/5 (100)	[28]
Giardia lamblia	GS/M (genotype B)	5×10^4	4/10 (40)	10/10 (100)	[30]
	Isr (genotype A)	5×10^4	0/5 (0)	0/5 (0)	[30]

In all bacterial challenge studies, the inocula were administered to fasting subjects with 2.0 g of $NaHCO_3$ (to neutralize gastric acid) except for reference [19], in which the inocula were administered in 45 mL of milk. *Giardia* trophozoites were administered directly into the proximal small by means of an intestinal tube (130-cm distance from the subject's mouth).

Abbreviations: CFU, colony-forming units; EPEC, enteropathogenic *Escherichia coli*; ETEC, enterotoxigenic *Escherichia coli*.

[a] Described as mild diarrhea.

[b] Described as severe diarrhea.

E74/68, *C. jejuni* A3249, and *G. lamblia* strain Isr rather than fully virulent strains. Until the specific virulence characteristics are identified that can differentiate highly pathogenic strains from strains that lack or have minimal pathogenicity, one cannot develop diagnostic tests to detect reliably the "true" pathogens which are expected to be found more often in cases of diarrhea, whereas the nonpathogenic varieties may be over-represented among isolates from controls.

For Some Pathogens, Clinical Illness May Require Interaction With a Second Pathogen, Whereas a Single Infection is Usually Asymptomatic

In the veterinary field, there are examples where, through a synergistic interaction, clinically overt or more severe diarrheal illness ensues when 2 specific enteric pathogens (such as ETEC and rotavirus) are present [31, 32]. In contrast, when the pathogens are present as single infections, diarrhea is milder or may not occur. Heretofore, examples of similar interactions of enteric pathogens in immunocompetent humans have not been convincingly described, but the possibility remains that they exist. Analyses of data from the Global Enteric Multicenter Study (GEMS) will offer the possibility of exploring that hypothesis.

HOST FACTORS

Host Susceptibility Factors

Host risk factors can play a critical role in the propensity to develop diarrheal illness or more severe illness following ingestion of a known enteropathogen. Many bacterial, viral, and

protozoal enteropathogens utilize molecules exposed on the surface of human intestinal cells as specific receptors to which they attach and initiate pathogenesis. The intestinal cell receptors include sugar moieties as well as proteins. Thus, susceptibility to infection and disease may be affected by the presence or absence of these receptors or the expression of variant receptors. Two striking examples of susceptibility based on genetic factors that involve blood group antigen expression are seen with cholera and norovirus infections. Human blood group antigens are expressed not only on erythrocytes but also on intestinal and other mucosal surfaces by genetically endowed persons ("secretors").

Cholera

Persons of blood group O and individuals with hypochlorhydria are much more prone to develop cholera gravis following the ingestion of a food or water vehicle containing *V. cholerae* O1 or *V. cholerae* O139. Blood group O has been recognized as a risk factor for cholera gravis both in epidemiologic field studies [33–36] and in volunteer challenge studies [37, 38]. In volunteer challenge studies a total purge of >5.0 liters of diarrheal stool is used as the definition of severe cholera and indicates a degree of purging that if not promptly and properly treated with aggressive rehydration would lead to cholera gravis, manifested by severe dehydration and hypovolemic shock. Another host risk factor for development of severe cholera is hypochlorhydria, with evidence deriving from clinical observations [39], epidemiologic studies, and volunteer challenge studies [40, 41].

Norovirus Gastroenteritis

Susceptibility to the Norwalk agent, the prototype GI-1 norovirus, is related to ABO blood group antigens. Volunteer studies showed that some individuals were highly resistant to Norwalk virus, whereas persons of blood group O exhibit increased risk of developing clinical illness upon exposure [42]. Norwalk virus binds to subjects whose intestinal secretions contain blood group O antigen H type 1 [43, 44], while norovirus GII-3 and GII-4 bind to cells of individuals who secrete blood group antigen A. Human hosts with null mutations of the gene encoding FUT2, the fucosyltransferase that determines secretor status, cannot synthesize ABH blood antigens in secretions. Such nonsecretors are in general not susceptible to norovirus disease [45], although recent epidemiologic studies suggest that some norovirus GII viruses can infect and cause disease even in nonsecretors [46, 47].

Other Nonspecific Host Factors That Affect Resistance to Diarrheal Pathogens

Various nonspecific but highly functional barriers protect the human intestine by impeding an enteric pathogen's ability to complete its pathogenesis that would otherwise result in clinical diarrheal illness [48]. One consequence of these barriers remaining intact is that the pathogen may end up colonizing the human intestine for a variable (short or long) period of time without causing overt diarrhea; this may explain some randomly selected matched control subjects in case/control studies who harbor pathogens in the absence of diarrhea. Barriers that a diarrheal pathogen must overcome include the intestinal microbiota (normal flora), the mucus layer, the epithelial cell layer, and various innate immune responses. These will be briefly mentioned in the ensuing paragraphs and recent reviews will be cited, should readers wish to delve deeper into these topics.

Intestinal Microbiota

The intestinal microbiota refers to the complex ecosystem of resident microorganisms (overwhelmingly either strict or facultative anaerobic bacteria) found in the mucus layer along the mucosal surface; enormous numbers (approximately $10^{12–14}$) of bacteria are found in the colon and terminal ileum [49]. In addition to performing symbiotic physiological functions for the host (eg, assisting in digestion, producing vitamin K and biotin, and promoting maturation of the mucosal immune system) [49–54], the microbiota constitute a formidable barrier that confronts pathogens [49–51, 54]. Besides competition for attachment sites on the epithelial surface and for nutrients, the end products of sugars metabolized by resident flora include short-chain fatty acids (eg, lactic, butyric, propionic) and other substances that are highly inhibitory for many bacterial enteropathogens such as *V. cholerae* O1 [55], *Salmonella*, and *Shigella* [56, 57].

Mucus Layer

The human intestine is covered by mucus, a product of goblet cells [58]. The mucus covering of the colon, composed of the mucin Muc2, is double layered, with the outer mucus layer being loosely adherent and replete with microbiota. In contrast, the inner mucus layer is highly adherent to the epithelium and is free of microorganisms [58, 59]. A healthy intact outer mucus layer constitutes a potent protective barrier that impedes enteropathogens. Beneath the mucus layer resides another defense barrier, the epithelial glycocalyx, consisting of diverse glycoproteins and glycolipids on the apical surface of enterocytes and colonocytes [60]. Both the mucus layer and the glycocalyx of the human intestine are continually replenished. The small intestine has only a single mucus layer. The mucus layer diminishes pathogen contact with the epithelium and carries bacteria distally [58].

Epithelial Cell Layer

The epithelial layer provides a 1-cell-thick physical barrier connected by tight junctions that separates pathogens in the

intestinal lumen from the lamina propria. In addition to the physical barrier, epithelial cells produce various antimicrobial peptides (defensins, cathelicidins, lysozyme, etc) [48]. Paneth cells, specialized secretory cells located in the crypts of the small intestine, are the primary source of the antimicrobial peptides [61, 62].

Various Innate Immune Responses

Epithelial cells and dendritic cells of the intestinal mucosa are replete with pathogen recognition receptors (PRRs) that detect the presence of pathogens and initiate a cascade of nonspecific innate immune responses that inhibit the pathogen. The PRRs include Toll-like receptors, nucleotide oligomerization domain–like receptors, retinoic-acid-inducible gene–like receptors, and the C-type lectin receptors [62].

Immune Status of the Host That Prevents Clinical Illness but Does Not Prevent Intestinal Colonization

Immune defenses such as intestinal secretory immunoglobulin A (sIgA) antibodies, breast milk sIgA antibodies or other nonspecific properties present in breast milk, or maternally derived serum immunoglobulin G (IgG) antibodies can prevent adherence of enteropathogens to enterocytes or mucosal invasion without killing the pathogen [63, 64]. Therefore, clinical illness is precluded, while still allowing asymptomatic intestinal carriage of the pathogen. The pathogens isolated from such asymptomatic individuals are nevertheless true pathogens. If these individuals are randomly selected healthy controls, they will be scored as control subjects carrying the pathogen(s) of interest. Below, several examples are given to illustrate these points.

Mucosal Immunity

The phenomenon of mucosal immunity providing clinical protection while still allowing asymptomatic excretion of pathogen is best illustrated with observations made in volunteer studies. North American volunteers who were vaccinated with a high dose (5×10^{10} colony-forming units [CFU]) of ETEC strain E1392-75-2A (O6:H16, LT/ST, CS1, CS3) mounted strong sIgA anti-CS1 and -CS3 antibody responses detected in jejunal fluids [65]. When 12 of these volunteers were challenged 1 month later with 5×10^8 CFU of wild-type strain E24377A (O139:H28, LT/ST, CS1,CS3), only 3 of 12 subjects developed diarrhea vs 6 of 6 unimmunized control subjects (75% vaccine efficacy; $P = .009$) [65]. An innovative facet of this study was the collection of jejunal fluids from the challenged vaccinees and control volunteers during late incubation and early in clinical illness to determine the presence and load of E7946 ETEC organisms in the proximal small intestine, the critical site of host–pathogen interaction. It is in the proximal small intestine that ETEC attaches to enterocytes by means of colonization factors and elaborate enterotoxins that culminate in diarrhea; stool culture positivity was also monitored. All 18 challenged subjects had positive stool cultures for the wild-type challenge organism, and all 6 controls had positive jejunal fluid cultures (with a mean of 7×10^3 CFU/mL). In contrast, only 1 vaccinee had a positive jejunal fluid culture following challenge ($P < .004$) and the colony count was only 10 CFU/mL [65–68]. Thus, in endemic areas where individuals are repetitively exposed to ETEC, individuals who have antiadhesin immunity in the proximal small intestine may be protected from ETEC diarrhea but may excrete the ETEC organisms in their stools.

Further observations supporting this phenomenon were made with infection-derived immunity to wild-type ETEC. Ten of 17 adult community volunteers developed watery diarrhea following ingestion of a dose of either 10^6 or 10^8 CFU of ETEC strain B7A with $NaHCO_3$ buffer [20] (Table 1). Eight of the 10 subjects who developed ETEC diarrhea were rechallenged 2 months later with 10^8 CFU (with buffer), along with 12 naive control subjects. Diarrhea developed in 7 of 12 controls but in only 1 of the 8 rechallenged "veterans" (75% efficacy, $P = .05$). Despite a significantly lower diarrhea attack rate, all 8 rechallenged veterans as well as all 12 controls had positive stool cultures for the ETEC challenge strain. A similar observation was also made during rechallenge studies with *Shigella flexneri* 2a [69]. A level of 70% clinical protection from prior clinical shigellosis was observed upon rechallenge, but all protected individuals shed *Shigella*, as did all naive controls. One must assume that a similar phenomenon of asymptomatic excretion among clinically protected persons living in ETEC and *Shigella*-endemic areas also occurs. If such individuals without diarrhea are randomly selected to serve as controls at a point when they are asymptomatically excreting ETEC, they will appear as culture-positive controls.

Breast Milk

Breastfeeding can protect infants and toddlers from developing more severe forms of diarrhea or even diarrhea at all [70, 71], without preventing intestinal colonization. Protection may be mediated by specific anti-pathogen sIgA antibodies in breast milk [72, 73] or by known nonspecific mechanisms such as lactoferrin [74, 75] and enterotoxin-binding oligosaccharides [76].

Transplacental Transfer of Maternal Antibodies

High titers of IgG maternal antibody against certain enteropathogens transferred transplacentally may prevent young infants from developing more severe forms of clinical illness infection or severe diarrheal disease until the titers wane [77, 78]. Because young infants in developing countries are also breastfed, it is challenging methodologically to isolate the relative contributions to protection that each of these confers.

Environmental Enteropathy

The syndrome of environmental enteropathy characterized by low-grade intestinal inflammation, blunted villi, increased numbers of intraepithelial and lamina propria lymphocytes, and proximal small bowel bacterial overgrowth is evident in a notable proportion of toddlers and preschool-aged children living in underprivileged conditions in developing countries [79–81]. The gut mucosa of these children is believed to have chronic activation of the innate immune system. In such children the ingestion of inocula that might be sufficient to cause diarrheal illness in a child without environmental enteropathy may be diminished by innate defenses such that colonization occurs but clinical disease does not. Environmental enteropathy may also play a role in diminishing the immune response of young children in developing countries to oral vaccines [81].

The Control Subject Is Incubating the Disease

The isolation of an enteropathogen from a control subject without diarrhea may in fact simply reflect identification of a recently exposed susceptible subject who is incubating the infection and will in 1 or more days develop diarrhea.

ENVIRONMENTAL FACTORS

Ingestion of an Inoculum Sufficient to Cause Subclinical Infection but Not Clinical Illness in a Susceptible Host

The presence of the pathogen in the stool of a healthy individual without diarrhea may reflect the recent ingestion of an inoculum too small to cause disease in an otherwise susceptible host; that is, if that individual had ingested a larger inoculum, diarrhea would have occurred. This may be particularly relevant for pathogens such as ETEC and *Salmonella* that are typically transmitted by food vehicles and that exhibit a clear dose-response curve (Table 1).

Ingestion of Host-Restricted Animal Pathogens

Porcine ETEC strain 263 causes severe dehydrating diarrhea in susceptible piglets. Following ingestion of 10^{10} CFU of this strain by adult volunteers, the strain was excreted but no subjects developed diarrhea. This is because the fimbrial colonization factor of this strain is specific for pigs but humans lack the receptors for attachment of the porcine fimbriae. In developing country niches where humans and animals such as pigs and bovines share close quarters, ingestion of animal ETEC incapable of causing human disease may be a common event. If animal ETEC is detected in a control subject without diarrhea by testing colonies for LT and ST and the colonies are not further characterized, they will be scored as ETEC.

OTHER FACTORS

Diagnostic Tests Vary Greatly in Their Sensitivity

Some diagnostic tests for enteropathogens, particularly molecular-based assays, may be so sensitive that they detect the passage through the gut of minute inocula of ingested pathogens that are insufficient to cause diarrhea. The peculiarities of different microbiological assays, including on detection of pathogens in control subjects, are discussed in the article by Robins-Browne and Levine in this supplement.

Disruption of the Intestinal Microbiome

Oral antibiotic use is promiscuous in developing countries and can alter the normal flora to render a human host susceptible to full-blown clinical infection, whereas in the absence of antibiotics, that host's unaltered flora might have interrupted the progression to diarrhea [82, 83]. Similarly, diet can markedly affect the composition of the microbiota [84].

Micronutrient Deficiency

Deficiency of zinc and vitamin A can increase the propensity of a child to develop clinically overt or more severe diarrheal illness following the ingestion of enteropathogens [84]. Conversely, pediatric subjects who do not manifest micronutrient deficiencies may be more likely to respond to the ingestion of enteropathogens by successfully limiting the infection to a subclinical state.

DISCUSSION

With modern, highly sensitive microbiologic methods and tests for pathogens that were unrecognized just a few decades ago, a wide array of enteropathogens can be recovered from cases of diarrhea in the GEMS. Indeed, the vast majority of GEMS patients with diarrhea can be expected to yield 1 or more possible etiologic agents. However, because of the pervasive fecal (human and animal) contamination that constitutes the underprivileged environment in which many young children are living in developing countries, facile transmission of pathogens readily occurs. It is therefore also imperative to assess the prevalence of various enteropathogens among appropriately selected subjects without diarrhea (ie, among matched controls). In a project such as GEMS, one expects to find a proportion of controls asymptomatically excreting known enteric pathogens. In this article we have attempted to review a series of plausible explanations for why healthy subjects without diarrhea may be excreting enteropathogens. To the best of our knowledge, this is the first time that these scenarios have been presented in a comprehensive way and from this perspective. Analyses of the GEMS epidemiologic, clinical, and microbiologic data in conjunction with detailed

characterization of specimens in the GEMS repository will allow us to address many of the hypotheses and commentaries raised in this review.

Notes

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol **1985**; 122:904–14.
2. Black RE, Lopez de Romana G, Brown KH, Bravo N, Bazalar OG, Kanashiro HC. Incidence and etiology of infantile diarrhea and major routes of transmission in Huascar, Peru. Am J Epidemiol **1989**; 129:785–99.
3. Steinsland H, Valentiner-Branth P, Perch M, et al. Enterotoxigenic *Escherichia coli* infections and diarrhea in a cohort of young children in Guinea-Bissau. J Infect Dis **2002**; 186:1740–7.
4. Hasan KZ, Pathela P, Alam K, et al. Aetiology of diarrhoea in a birth cohort of children aged 0–2 year(s) in rural Mirzapur, Bangladesh. J Health Popul Nutr **2006**; 24:25–35.
5. Ferreccio C, Prado V, Ojeda A, et al. Epidemiologic patterns of acute diarrhea and endemic *Shigella* infections in a poor periurban setting in Santiago, Chile. Am J Epidemiol **1991**; 134:614–27.
6. Levine MM, Ferreccio C, Prado V, et al. Epidemiologic studies of *Escherichia coli* infections in a low socioeconomic level periurban community in Santiago, Chile. Am J Epidemiol **1993**; 138:849–69.
7. Buchwald DS, Blaser MJ. A review of human salmonellosis: II. Duration of excretion following infection with nontyphi *Salmonella*. Rev Infect Dis **1984**; 6:345–56.
8. Smith ER, Badley BW. Treatment of *Salmonella* enteritis and its effect on the carrier state. Can Med Assoc J **1971**; 104:1004–6.
9. Tribble DR, Baqar S, Scott DA, et al. Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans. Infect Immun **2010**; 78:1750–9.
10. Salazar-Lindo E, Sack RB, Chea-Woo E, et al. Early treatment with erythromycin of *Campylobacter jejuni*-associated dysentery in children. J Pediatr **1986**; 109:355–60.
11. Robins-Browne RM, Mackenjee MK, Bodasing MN, Coovadia HM. Treatment of *Campylobacter*-associated enteritis with erythromycin. Am J Dis Child **1983**; 137:282–5.
12. Porter IA, Reid TM. A milk-borne outbreak of *Campylobacter* infection. J Hyg (Lond) **1980**; 84:415–9.
13. Partridge DG, Evans CM, Raza M, Kudesia G, Parsons HK. Lessons from a large norovirus outbreak: impact of viral load, patient age and ward design on duration of symptoms and shedding and likelihood of transmission. J Hosp Infect **2012**; 81:25–30.
14. Aoki Y, Suto A, Mizuta K, Ahiko T, Osaka K, Matsuzaki Y. Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients. J Hosp Infect **2010**; 75:42–6.
15. Henke-Gendo C, Harste G, Juergens-Saathoff B, Mattner F, Deppe H, Heim A. New real-time PCR detects prolonged norovirus excretion in highly immunosuppressed patients and children. J Clin Microbiol **2009**; 47:2855–62.
16. Atmar RL, Opekun AR, Gilger MA, et al. Norwalk virus shedding after experimental human infection. Emerg Infect Dis **2008**; 14:1553–7.
17. Levine MM, DuPont HL, Khodabandelou M, Hornick RB. Long-term *Shigella*-carrier state. N Engl J Med **1973**; 288:1169–71.
18. Levine M, Kaper J, Black RE, Clements M. New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development. Microbiol Rev **1983**; 47:510–50.
19. DuPont HL, Formal SB, Hornick RB, et al. Pathogenesis of *Escherichia coli* diarrhea. N Engl J Med **1971**; 285:1–9.
20. Levine MM, Nalin DR, Hoover DL, Bergquist EJ, Hornick RB, Young CR. Immunity to enterotoxigenic *Escherichia coli*. Infect Immun **1979**; 23:729–36.
21. Levine MM, Caplan ES, Waterman D, Cash RA, Hornick RB, Snyder MJ. Diarrhea caused by *Escherichia coli* that produce only heat- stable enterotoxin. Infect Immun **1977**; 17:78–82.
22. Lumish RM, Ryder RW, Anderson DC, Wells JG, Puhr ND. Heat-labile enterotoxigenic *Escherichia coli* induced diarrhea aboard a Miami-based cruise ship. Am J Epidemiol **1980**; 111:432–6.
23. Evans DG, Satterwhite TK, Evans DJ Jr, DuPont HL. Differences in serological responses and excretion patterns of volunteers challenged with enterotoxigenic *Escherichia coli* with and without the colonization factor antigen. Infect Immun **1978**; 19:883–8.
24. Satterwhite TK, Evans DG, DuPont HL, Evans DJ Jr. Role of *Escherichia coli* colonisation factor antigen in acute diarrhoea. Lancet **1978**; 2:181–4.
25. Tacket CO, Losonsky G, Link H, et al. Protection by milk immunoglobulin concentrate against oral challenge with enterotoxigenic *Escherichia coli*. N Engl J Med **1988**; 318:1240–3.
26. Levine MM, Black RE, Clements ML, Nalin DR, Cisneros L, Finkelstein RA. Volunteer studies in development of vaccines against cholera and enterotoxigenic *Escherichia coli*: a review. In: Holme T, Holmgren J, Merson MH, Mollby R, eds. Acute enteric infections in children. New prospects for treatment and prevention. Amsterdam: Elsevier/North-Holland Biomedical Press, **1981**:443–59.
27. Levine MM, Bergquist EJ, Nalin DR, et al. *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet **1978**; 1:1119–22.
28. Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. Experimental *Campylobacter jejuni* infection in humans. J Infect Dis **1988**; 157:472–9.
29. Nash TE, Herrington DA, Levine MM, Conrad JT, Merritt JW Jr. Antigenic variation of *Giardia lamblia* in experimental human infections. J Immunol **1990**; 144:4362–9.
30. Nash TE, Herrington DA, Losonsky GA, Levine MM. Experimental human infections with *Giardia lamblia*. J Infect Dis **1987**; 156:974–84.
31. Neog BK, Barman NN, Bora DP, Dey SC, Chakraborty A. Experimental infection of pigs with group A rotavirus and enterotoxigenic *Escherichia coli* in India: gross, histopathological and immunopathological study. Vet Ital **2011**; 47:117–28.
32. Tzipori S, Chandler D, Makin T, Smith M. *Escherichia coli* and rotavirus infections in four-week-old gnotobiotic piglets fed milk or dry food. Aust Vet J **1980**; 56:279–84.
33. Glass RI, Becker S, Huq I, et al. Endemic cholera in rural Bangladesh, 1966–1980. Am J Epidemiol **1982**; 116:959–70.
34. Barua D. ABO blood groups and cholera. Ann Hum Biol **1977**; 4:489–92.
35. Harris JB, Khan AI, LaRocque RC, et al. Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. Infect Immun **2005**; 73:7422–7.
36. Faruque AS, Mahalanabis D, Hoque SS, Albert MJ. The relationship between ABO blood groups and susceptibility to diarrhea due to *Vibrio cholerae* 0139. Clin Infect Dis **1994**; 18:827–8.
37. Tacket CO, Losonsky G, Nataro JP, et al. Extension of the volunteer challenge model to study South American cholera in a population of volunteers predominantly with blood group antigen O. Trans R Soc Trop Med Hyg **1995**; 89:75–7.
38. Tacket CO, Cohen MB, Wasserman SS, et al. Randomized, double-blind, placebo-controlled, multicentered trial of the efficacy of a single

dose of live oral cholera vaccine CVD 103-HgR in preventing cholera following challenge with vibrio cholerae O1 El Tor Inaba three months after vaccination. Infect Immun **1999**; 67:6341–5.
39. Gitelson S. Gastrectomy, achlorhydria and cholera. Isr J Med Sci **1971**; 7:663–7.
40. Nalin DR, Levine MM, Rhead J, et al. Cannabis, hypochlorhydria, and cholera. Lancet **1978**; 2:859–62.
41. Nalin DR, Levine RJ, Levine MM, et al. Cholera, non-vibrio cholera, and stomach acid. Lancet **1978**; 2:856–9.
42. Hutson AM, Atmar RL, Graham DY, Estes MK. Norwalk virus infection and disease is associated with ABO histo-blood group type. J Infect Dis **2002**; 185:1335–7.
43. Flores J, Okhuysen PC. Genetics of susceptibility to infection with enteric pathogens. Curr Opin Infect Dis **2009**; 22:471–6.
44. Hutson AM, Atmar RL, Marcus DM, Estes MK. Norwalk virus-like particle hemagglutination by binding to h histo-blood group antigens. J Virol **2003**; 77:405–15.
45. Lindesmith L, Moe C, Marionneau S, et al. Human susceptibility and resistance to Norwalk virus infection. Nat Med **2003**; 9:548–53.
46. Carlsson B, Kindberg E, Buesa J, et al. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 norovirus infection. PLoS One **2009**; 4:e5593.
47. Halperin T, Vennema H, Koopmans M, et al. No association between histo-blood group antigens and susceptibility to clinical infections with genogroup II norovirus. J Infect Dis **2008**; 197:63–5.
48. Gill N, Wlodarska M, Finlay BB. Roadblocks in the gut: barriers to enteric infection. Cell Microbiol **2011**; 13:660–9.
49. Lievin-Le Moal V, Servin AL. The front line of enteric host defense against unwelcome intrusion of harmful microorganisms: mucins, antimicrobial peptides, and microbiota. Clin Microbiol Rev **2006**; 19:315–37.
50. Yu LC, Wang JT, Wei SC, Ni YH. Host-microbial interactions and regulation of intestinal epithelial barrier function: from physiology to pathology. World J Gastrointest Pathophysiol **2012**; 3:27–43.
51. Stecher B, Hardt WD. Mechanisms controlling pathogen colonization of the gut. Curr Opin Microbiol **2011**; 14:82–91.
52. Ivanov II, Littman DR. Modulation of immune homeostasis by commensal bacteria. Curr Opin Microbiol **2011**; 14:106–14.
53. Wardwell LH, Huttenhower C, Garrett WS. Current concepts of the intestinal microbiota and the pathogenesis of infection. Curr Infect Dis Rep **2011**; 13:28–34.
54. Guarner F, Malagelada JR. Gut flora in health and disease. Lancet **2003**; 361:512–9.
55. Silva SH, Vieira EC, Dias RS, Nicoli JR. Antagonism against *Vibrio cholerae* by diffusible substances produced by bacterial components of the human faecal microbiota. J Med Microbiol **2001**; 50:161–4.
56. Maier BR, Onderdonk AB, Baskett RC, Hentges DJ. *Shigella*, indigenous flora interactions in mice. Am J Clin Nutr **1972**; 25:1433–40.
57. Levine MM, Hornick RB. Lactulose therapy in *Shigella* carrier state and acute dysentery. Antimicrob Agents Chemother **1975**; 8:581–4.
58. Hansson GC. Role of mucus layers in gut infection and inflammation. Curr Opin Microbiol **2012**; 15:57–62.
59. Johansson ME, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci U S A **2008**; 105:15064–9.
60. Moran AP, Gupta A, Joshi L. Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. Gut **2011**; 60:1412–25.
61. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nat Rev Microbiol **2011**; 9:356–68.
62. Kinnebrew MA, Pamer EG. Innate immune signaling in defense against intestinal microbes. Immunol Rev **2012**; 245:113–31.
63. Willer EM, Lima RL, Giugliano LG. In vitro adhesion and invasion inhibition of *Shigella dysenteriae*, *Shigella flexneri* and *Shigella sonnei* clinical strains by human milk proteins. BMC Microbiol **2004**; 4:18.
64. Hayani KC, Guerrero ML, Morrow AL, et al. Concentration of milk secretory immunoglobulin A against *Shigella* virulence plasmid-associated antigens as a predictor of symptom status in *Shigella*-infected breast-fed infants. J Pediatr **1992**; 121:852–6.
65. Levine MM, Morris JG, Losonsky G, Boedeker E, Rowe B. Fimbriae (pili) adhesins as vaccines. In: Lark DL, Normark S, Uhlin BE, Wolf-Watz H, eds. Protein-carbohydrate interactions in biological systems: the molecular biology of microbial pathogenicity. London: Academic Press, **1986**:143–5.
66. Levine MM, Ristaino P, Marley G, et al. Coli surface antigens 1 and 3 of colonization factor antigen II- positive enterotoxigenic *Escherichia coli*: morphology, purification, and immune responses in humans. Infect Immun **1984**; 44:409–20.
67. Levine MM, Black RE, Clements ML, et al. Prevention of enterotoxigenic *Escherichia coli* diarrheal infection by vaccines that stimulate antiadhesion (antipili) immunity. In: Boedeker EC, ed. Attachment of organisms to the gut mucosa. Boca Raton, FL: CRC Press, **1984**:223–44.
68. Levine MM. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J Infect Dis **1987**; 155:377–89.
69. Kotloff KL, Nataro JP, Losonsky GA, et al. A modified *Shigella* volunteer challenge model in which the inoculum is administered with bicarbonate buffer: clinical experience and implications for *Shigella* infectivity. Vaccine **1995**; 13:1488–94.
70. Ahmed F, Clemens JD, Rao MR, Sack DA, Khan MR, Haque E. Community-based evaluation of the effect of breast-feeding on the risk of microbiologically confirmed or clinically presumptive shigellosis in Bangladeshi children. Pediatrics **1992**; 90:406–11.
71. Clemens JD, Stanton B, Stoll B, Shahid NS, Banu H, Chowdhury AK. Breast feeding as a determinant of severity in shigellosis. Evidence for protection throughout the first three years of life in Bangladeshi children. Am J Epidemiol **1986**; 123:710–20.
72. Brandtzaeg P. The mucosal immune system and its integration with the mammary glands. J Pediatr **2010**; 156(2 suppl):S8–15.
73. Glass RI, Svennerholm AM, Stoll BJ, et al. Protection against cholera in breast-fed children by antibodies in breast milk. N Engl J Med **1983**; 308:1389–92.
74. Giugliano LG, Ribeiro ST, Vainstein MH, Ulhoa CJ. Free secretory component and lactoferrin of human milk inhibit the adhesion of enterotoxigenic *Escherichia coli*. J Med Microbiol **1995**; 42:3–9.
75. Walker A. Breast milk as the gold standard for protective nutrients. J Pediatr **2010**; 156(2 suppl):S3–7.
76. Kunz C, Rudloff S. Potential anti-inflammatory and anti-infectious effects of human milk oligosaccharides. Adv Exp Med Biol **2008**; 606:455–65.
77. Ramachandran M, Vij A, Kumar R, et al. Lack of maternal antibodies to P serotypes may predispose neonates to infections with unusual rotavirus strains. Clin Diagn Lab Immunol **1998**; 5: 527–30.
78. Mutanda LN. Epidemiology of acute gastroenteritis in early childhood in Kenya. V. An inverse relationship between the peak age-incidence and the waning of rotavirus maternal antibodies. East Afr Med J **1980**; 57:545–8.
79. Fagundes Neto U, Martins MC, Lima FL, Patricio FR, Toledo MR. Asymptomatic environmental enteropathy among slum-dwelling infants. J Am Coll Nutr **1994**; 13:51–6.
80. dos Reis JC, de Morais MB, Oliva CA, Fagundes-Neto U. Breath hydrogen test in the diagnosis of environmental enteropathy in children living in an urban slum. Dig Dis Sci **2007**; 52:1253–8.

81. Levine MM. Immunogenicity and efficacy of oral vaccines in developing countries: lessons from a live cholera vaccine. BMC Biol **2010**; 8:129.
82. Looft T, Allen HK. Collateral effects of antibiotics on mammalian gut microbiomes. Gut Microbes **2012**; 3:1–5.
83. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol **2008**; 6:e280.
84. Kelly P. Nutrition, intestinal defence and the microbiome. Proc Nutr Soc **2010**; 69:261–8.
85. Levine MM. Immunity to cholera as evaluated in volunteers. In: Ouchterlony O, Holmgren J, eds. Cholera and related diarrheas. Basel: Karger, **1980**:195–203.
86. Levine MM, Black RE, Clements ML, et al. The pathogenicity of nonenterotoxigenic *Vibrio cholerae* serogroup O1 biotype El Tor isolated from sewage water in Brazil. J Infect Dis **1982**; 145: 296–9.
87. Donnenberg MS, Tacket CO, James SP, et al. Role of the eaeA gene in experimental enteropathogenic *Escherichia coli* infection. J Clin Invest **1993**; 92:1412–7.

SUPPLEMENT ARTICLE

Laboratory Diagnostic Challenges in Case/Control Studies of Diarrhea in Developing Countries

Roy M. Robins-Browne[1,2] and Myron M. Levine[3]

[1]Department of Microbiology and Immunology, The University of Melbourne, [2]Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia; and [3]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

Case/control studies of acute infectious diarrhea require accurate and dependable laboratory tests to detect pathogens in samples from both symptomatic patients and healthy control subjects. The methods used to detect these pathogens have usually been evaluated on patient samples only, and their performance on samples from control subjects is mostly unknown. Because many pathogens occur at a high overall frequency in developing countries and thus may be present in a notable proportion of control subjects as well as patients, the relative ability of a diagnostic test to detect these pathogens in diarrheic and normal stools can have a profound effect on the interpretation of case/control data.

The laboratory procedures used to detect etiologic agents in patients with acute infectious diarrhea are constantly evolving. Nevertheless, the principles underlying the performance and interpretation of these procedures are well established. As with clinical diagnostic microbiology in general, the choice of the tests used to detect a possible etiologic agent is determined by clinical relevance, practicability, and cost. For the most part, this approach is satisfactory, although in some cases no etiologic agent is identified. This may be because the diarrhea is not infectious in origin, or because a particular agent is not identified either because it was not sought or because the procedures used to detect it were not sufficiently sensitive and gave a false-negative result. On the other hand, when a single pathogen is found, interpretation of the results is straightforward, insofar as the pathogen is usually assumed to be responsible for the patient's symptoms.

In case/control studies, in addition to investigating diarrheic samples for pathogens, we undertake the far less familiar task of investigating feces or rectal swabs from subjects without diarrhea. The detection of a pathogen in these individuals indicates asymptomatic carriage, the possible reasons for which are discussed in the accompanying article by Levine and Robins-Browne in this supplement.

Analysis of the outcome of a case/control study involves comparing the frequencies of the detection of pathogens in cases and control subjects, which are used to determine an odds ratio (OR). As the OR indicates the strength of the association between a pathogen and the occurrence of diarrhea, it is used as a measure of the relative pathogenicity of different pathogens. The OR is also one factor in the equation utilized to calculate attributable fraction, which provides an estimate of the relative contribution of the pathogen(s) of interest to the diarrheal disease burden (see the article by Blackwelder et al in this supplement). Because the frequency of detection of pathogens in control subjects can have a profound effect on the interpretation of case/control data, it is essential to understand the performance of laboratory tests in samples from subjects without diarrhea, as well as those from patients.

Correspondence: Roy M. Robins-Browne, MB, BCh, PhD, Dept of Microbiology and Immunology, Grattan St, The University of Melbourne, Victoria 3010, Australia (r.browne@unimelb.edu.au).

Clinical Infectious Diseases 2012;55(S4):S312–16

DOI: 10.1093/cid/cis756

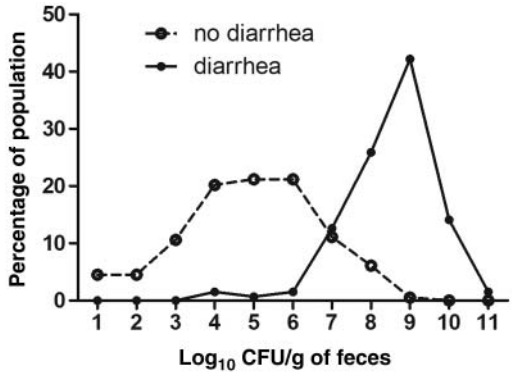

Figure 1. Frequency distribution of the number of the colony-forming units from 198 fecal samples from rabbits without diarrhea, and 135 samples from rabbits with diarrhea caused by infection with 1 of 3 different rabbit-specific enteropathogenic *Escherichia coli* strains. Data were obtained from quantitative cultures of samples on selective agar containing antibiotics to which the infecting strains were resistant. The data show that feces from rabbits without diarrhea contain significantly lower numbers of bacteria than those with diarrhea. Rabbits without diarrhea included similar numbers of animals with subclinical infection and animals sampled during the incubation or convalescent periods of symptomatic infection. Abbreviation: CFU, colon-forming units.

Two key factors that govern the usefulness and reliability of a laboratory test are its sensitivity and specificity. For almost all tests, these parameters are first determined by using "spiked" samples and then evaluated under field conditions, often in comparison with other tests. A test that is highly sensitive and specific will reliably detect a pathogen in cases with few false-negative or false-positive results. However, diagnostic tests are seldom evaluated in control subjects (ie, individuals without symptoms). In industrialized countries this is seldom an issue because the prevalence of most pathogens in healthy subjects is low. In developing countries, however, where sanitation is poor and exposure to contaminated food and water is virtually a daily norm, intestinal pathogens circulate at high frequency and children are liable to become repeatedly infected with them. Accordingly, endemic pathogens will be present at a far higher frequency overall than in industrialized countries. The ability to detect these pathogens in control subjects will differ according to the sensitivity of the diagnostic procedures that are used.

Table 1. Results of Hypothetical Case/Control Studies Where the Frequency of a Pathogen in Patients Is Fixed at 25% and Its Frequency in Control Subjects Ranges From 0% to 25%

Frequency in 100 Patients	Frequency in 100 Control Subjects	Odds Ratio	*P* Value[a]
25	0	Infinity	<.0001
25	5	6.3	.001
25	10	3	.009
25	15	1.9	.11
25	20	1.3	.5
25	25	1	1

[a] Fisher exact test, 2-tailed.

Most comprehensive case/control studies of diarrhea include culture for bacterial pathogens, such as *Salmonella* species, *Shigella* species, and *Escherichia coli*. Culture of fecal samples has an intrinsically low sensitivity to detect pathogens, especially in individuals without diarrhea, because the complex microbiota of healthy individuals makes it difficult to detect a pathogen among the high background "noise." The fact that culture of feces is a useful diagnostic procedure despite its low sensitivity can be explained partly by the fact that in patients with diarrhea, the pathogen is generally excreted in far higher numbers and makes up a much greater proportion of the cultivable microbiota than in healthy subjects who are asymptomatic carriers of the same pathogen. The odds of finding a pathogen in both cases and controls can be considerably improved by using selective media with or without prior enrichment. The use of such media has revolutionized our understanding of the epidemiology of bacterial enteropathogens such as *Campylobacter jejuni* in developing and industrialized countries [1, 2].

To illustrate these points, we will use data from studies we have undertaken with an animal model of diarrhea caused by a subtype of enteropathogenic *E. coli*, known as rabbit-specific enteropathogenic *E. coli* (REPEC). Infection of infant rabbits with REPEC closely parallels infection of human infants with human-specific enteropathogenic *E. coli* (EPEC) in terms of age-related susceptibility, clinical presentation, and associated intestinal pathology [3, 4].While establishing this model at the University of Melbourne and determining the median infectious dose of different REPEC strains, the natural course of infection was charted by observing rabbits for symptoms of diarrhea and correlating this with quantitative culture of REPEC on selective media containing antibiotics to which the challenge strains were resistant. Detailed descriptions of our methods have been published previously [5].

To investigate the hypothesis that bacteria are present in greater numbers (and therefore more easily detected) in cases with diarrhea than in control subjects, we reanalyzed published and unpublished data from experiments in which we infected rabbits with 1 of 3 different wild-type strains of REPEC of differing virulence. Quantitative culture of fecal samples from these animals indicated that rabbits with diarrhea excrete significantly more bacteria (1.1×10^8 colony-forming units [CFU] per gram of feces [mean]; 1.6×10^8 CFU [median]) than rabbits without symptoms (2.8×10^4 CFU

Table 2. Results of Hypothetical Case/Control Studies Where the Frequency of a Pathogen in Patients Ranges From 10% to 30%, and the Difference in Its Frequency in Patients and Control Subjects Is Fixed at 10%

Frequency in 100 Patients	Frequency in 100 Control Subjects	Odds Ratio	*P* Value[a]
10	0	Infinity	.001
15	5	3.4	.03
20	10	2.3	.07
25	15	1.9	.11
30	20	1.7	.14

[a] Fisher exact test, 2-tailed.

[mean]; 3.9×10^4 CFU [median]; $P < .0001$) (Figure 1). Our data indicated that a test with a detection limit of 10^7 CFU per gram would be positive in 96% of cases of diarrhea and in 18% of infected, but asymptomatic, individuals (Figure 1). In contrast, a test with a detection limit of 10^4 CFU per gram would be positive in 100% of cases and 80% of infected controls. This analysis exemplifies how increasing the sensitivity of the test can improve detection limits disproportionately in control samples compared with samples from patients. The influence this may have on the interpretation of hypothetical case/control data is shown in Tables 1 and 2. Our data suggest that as test sensitivity increases, quantitative assays may be useful in distinguishing clinical from subclinical infection. This suggestion has been borne out by a recent report by Barletta et al [6], who found that the use of quantitative polymerase chain reaction (PCR) to diagnose EPEC infection in children in an endemic setting yielded higher values in patients than in subjects without diarrhea.

A striking example of how improved detection of intestinal pathogens can influence data obtained from a case/control study comes from the comprehensive English Infectious Intestinal Disease Study (1993–1996). By using PCR for 8 groups of pathogens to investigate 4627 archived fecal samples from 2422 cases and 2205 controls in the original study (which did not use PCR-based detection of enteropathogens), Amar et al [7] increased the detection rate of at least 1 agent (or toxin) from 53% in the original study to 75% in cases, and from 19% to 42% in controls. Furthermore, the use of PCR-based diagnosis for 8 groups of pathogens increased the number of cases in whom >1 pathogen was detected from 272 to 993 (a 73% increase), and from 32 to 280 (a 89% increase) in controls. The greatest increase in detection rates that resulted from the use of PCR was for rotavirus and norovirus.

For example, in children aged <1 year the detection rate of rotavirus in cases went from 29 of 144 (20%) detected by enzyme immunoassay (EIA) and/or electron microscopy (EM) to 70 of 144 (49%) detected by PCR and EM. In controls the increase in diagnostic yield using PCR was even greater: from 3 of 183 (2%) to 53 of 183 (29%). In this age group, the OR before PCR was 15.1 (95% confidence interval [CI], 4.5–78.8); and with the PCR data included it was 2.3 (95% CI, 1.4–3.7). Despite the fall in OR, the attributable fraction (discussed in the article by Blackwelder et al in this supplement) increased from 19% using the original detection method to 28% using PCR. This can be explained by the significantly increased detection of rotavirus in cases. Also, in the original study, norovirus was detected by using EM in 14 of 144 (10%) cases <1 year old and in 2 of 183 (1%) age-matched controls (OR, 9.7 [95% CI, 2.2–89.3]). Together, however, EM and PCR revealed norovirus in 70 of 144 cases (49%) and 57 of 183 (31%) controls (OR, 2.1 [95% CI, 1.3–3.4]). In this case, the attributable fraction increased from 9% using the original method to 25% using more sensitive detection. These data, albeit from an industrialized country, clearly indicate how using a test with enhanced sensitivity can influence the major outcomes of a case/control study. Amar et al [7] also reported that quantitative PCR for norovirus may permit asymptomatic carriage to be distinguished from symptomatic infection, a finding that was subsequently confirmed by other researchers [8]. Similar findings linking the number of virus particles in feces to disease severity have also been reported for rotavirus [9, 10].

Although mucosal immunity will account for some instances of asymptomatic carriage of particular pathogens, the presence of mucosal antibodies may also interfere with the ability to detect pathogens when using EIA. In the case of rotavirus, for example, the most common method of diagnosis is a type of "sandwich" EIA, in which an immobilized antibody is used to capture a rotavirus antigen from feces, after which the captured antigen is revealed by using a second, labeled antibody. Tests of this type are capable of detecting between 10^5 and 10^6 rotavirus particles per milliliter [11]. However, the sensitivity of this assay may fall during the course of the illness, as patients develop immunity to rotavirus and secrete mucosal antibodies that coat the virus and interfere with its detection by EIA [11, 12]. By contrast, PCR using reverse transcription to amplify rotavirus RNA is able to detect as few as 1000 virus particles per milliliter [11], and is unaffected by mucosal immunity. In a study of children hospitalized for diarrhea with rotavirus, PCR-based diagnosis revealed that 11 of 37 (30%) children were still infected with rotavirus >3 weeks after hospitalization compared with only 2 of 37 (5%) when EIA was used to detect the virus [11]. These data indicate that the EIA for rotavirus is more likely to be positive in patients experiencing their first infection with a particular virus than in children who are convalescing from an acute infection or are reinfected with a strain of rotavirus they have encountered previously. By contrast, the EIA for *Giardia lamblia* is extremely sensitive and can be used to identify asymptomatic

carriers of this pathogen, for example, during the investigation of outbreaks of giardiasis in industrialized countries [13].

Apart from test sensitivity, another factor that may influence the comparison of laboratory data from cases and control subjects in case/control studies of diarrhea is the nature of the samples that are investigated. In an ideal case/control study, the diagnostic samples that are collected from patients and controls should be the same. In case/control studies of diarrhea, however, this is generally not the case, because in patients with diarrhea, especially watery diarrhea, much of the sample will have originated in the small intestine, and the normal microbiota of the large intestine will have been purged or significantly diluted, whereas in controls, the fecal samples or rectal swabs that are investigated will reflect the microbiota of the distal large intestine.

Although this difference may not matter in some instances, in others it could be important. For example, some bacteria only cause disease in the small intestine. An example is enterotoxigenic *E. coli* (ETEC), in which enterotoxins that are responsible for diarrhea act predominantly in the small intestine [14]. Furthermore, studies with EPEC infection in adult volunteers have shown that these bacteria are virulent only when ingested by mouth and not when they are inoculated directly into the large intestine [15]. This may be explained by the fact that environmental signals required to activate virulence gene expression are absent from the large intestine [16]. Given that *E. coli* is well adapted to persist in the large intestine as part of the normal microbiota, it is conceivable that strains of pathogenic *E. coli*, including ETEC, EPEC, and enteroaggregative *E. coli*, may colonize the large intestine of healthy people or convalescent patients and behave as nonpathogens, whereas the same bacteria isolated from the small intestine would be of considerable diagnostic significance. Similar circumstances may apply to other enteric pathogens that also differ in their ability to cause disease depending on their site of intestinal colonization. On the other hand, diarrheic stools may contain commensal microorganisms, which normally reside mainly in the proximal intestine and are not readily detectable in formed stools. In this case, the association of the agent with diarrhea may lead to the false conclusion that it is a causative agent.

As detailed in the article in this supplement by Nataro et al, the methods used to detect and identify pathogens in The Global Enteric Multicenter Study (GEMS) were state-of-the-art. Nevertheless, the relative ability of these methods to detect pathogens in cases and control subjects in developing countries is not known. For a thorough understanding of case/control data, especially when comparing the relative contribution of different pathogens to the overall burden of disease, we need a more thorough understanding of the performance of diagnostic procedures as used on samples from cases and controls. Some possible areas of further study could include the quantitative analysis of patients' samples (particularly when there is >1 pathogen), and examination of virulence gene expression to indicate if a putative etiologic agent is behaving as a pathogen or commensal.

Notes

Acknowledgments. We are grateful to Dr William Blackwelder and Dr Tamer Farag for helpful discussion and advice, and to Louise Adams for providing unpublished data.

Financial support. This research was supported by the Australian National Health and Medical Research Council and the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Blaser MJ, Taylor DN, Feldman RA. Epidemiology of *Campylobacter jejuni* infections. Epidemiol Rev **1983**; 5:157–76.
2. Patton CM, Mitchell SW, Potter ME, Kaufmann AF. Comparison of selective media for primary isolation of *Campylobacter fetus* subsp. *jejuni*. J Clin Microbiol **1981**; 13:326–30.
3. Clarke SC, Haigh RD, Freestone PP, Williams PH. Virulence of enteropathogenic *Escherichia coli*, a global pathogen. Clin Microbiol Rev **2003**; 16:365–78.
4. Milon A, Oswald E, De RJ. Rabbit EPEC: a model for the study of enteropathogenic *Escherichia coli*. Vet Res **1999**; 30:203–19.
5. Robins-Browne RM, Tokhi AM, Adams LM, Bennett-Wood V. Host specificity of enteropathogenic *Escherichia coli* from rabbits: lack of correlation between adherence in vitro and pathogenicity for laboratory animals. Infect Immun **1994**; 62:3329–36.
6. Barletta F, Ochoa TJ, Mercado E, et al. Quantitative real-time polymerase chain reaction for enteropathogenic *Escherichia coli*: a tool for investigation of asymptomatic versus symptomatic infections. Clin Infect Dis **2011**; 53:1223–9.
7. Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993–1996). Eur J Clin Microbiol Infect Dis **2007**; 26:311–23.
8. Phillips G, Lopman B, Tam CC, Iturriza-Gomara M, Brown D, Gray J. Diagnosing norovirus-associated infectious intestinal disease using viral load. BMC Infect Dis **2009**; 9:63.
9. Kang G, Iturriza-Gomara M, Wheeler JG, et al. Quantitation of group A rotavirus by real-time reverse-transcription-polymerase chain reaction: correlation with clinical severity in children in South India. J Med Virol **2004**; 73:118–22.
10. Phillips G, Lopman B, Tam CC, Iturriza-Gomara M, Brown D, Gray J. Diagnosing rotavirus A associated IID: using ELISA to identify a cut-off for real time RT-PCR. J Clin Virol **2009**; 44:242–5.
11. Richardson S, Grimwood K, Gorrell R, Palombo E, Barnes G, Bishop R. Extended excretion of rotavirus after severe diarrhoea in young children. Lancet **1998**; 351:1844–8.
12. Coulson BS, Grimwood K, Hudson IL, Barnes GL, Bishop RF. Role of coproantibody in clinical protection of children during reinfection with rotavirus. J Clin Microbiol **1992**; 30:1678–84.
13. Garcia LS, Shimizu RY. Evaluation of nine immunoassay kits (enzyme immunoassay and direct fluorescence) for detection of *Giardia lamblia* and *Cryptosporidium parvum* in human fecal specimens. J Clin Microbiol **1997**; 35:1526–9.

14. Turner SM, Scott-Tucker A, Cooper LM, Henderson IR. Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. FEMS Microbiol Lett **2006**; 263:10–20.
15. Koya G, Kosakai N, Kono M, Mori M, Fukasawa Y. Observations on the multiplication of *Escherichia coli* O-111 B4 in the intestinal tract of adult volunteers in feeding experiments: the intubation study with Miller-Abbott's double lumen tube. Jpn J Med Sci Biol **1954**; 7:197–203.
16. Yang J, Tauschek M, Robins-Browne RM. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. Trends Microbiol **2011**; 19:128–35.

SUPPLEMENT ARTICLE

Exploring Household Economic Impacts of Childhood Diarrheal Illnesses in 3 African Settings

Richard Rheingans,[1] Matt Kukla,[1] Richard A. Adegbola,[2,3] Debasish Saha,[2] Richard Omore,[4,5] Robert F. Breiman,[6] Samba O. Sow,[7] Uma Onwuchekwa,[7] Dilruba Nasrin,[8] Tamer H. Farag,[8] Karen L. Kotloff,[8] and Myron M. Levine[8]

[1]Department of Environmental and Global Health, University of Florida, Gainesville; [2]Medical Research Council Unit, Fajara and Basse Stations, The Gambia, West Africa; [3]GlaxoSmithKline Biologicals, Global Medical Affairs, Wavre, Belgium; [4]Kenya Medical Research Institute / Centers for Disease Control and Prevention (KEMRI/CDC), [5]Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, [6]Global Disease Detection Division, Kenya Office of the US Centers for Disease Control and Prevention, Nairobi; [7]Center for Vaccine Development–Mali, Bamako; and [8]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

Beyond the morbidity and mortality burden of childhood diarrhea in sub-Saharan African are significant economic costs to affected households. Using survey data from 3 of the 4 sites in sub-Saharan Africa (Gambia, Kenya, Mali) participating in the Global Enteric Multicenter Study (GEMS), we estimated the direct medical, direct nonmedical, and indirect (productivity losses) costs borne by households due to diarrhea in young children. Mean cost per episode was $2.63 in Gambia, $6.24 in Kenya, and $4.11 in Mali. Direct medical costs accounted for less than half of these costs. Mean costs understate the distribution of costs, with 10% of cases exceeding $6.50, $11.05, and $13.84 in Gambia, Kenya, and Mali. In all countries there was a trend toward lower costs among poorer households and in 2 of the countries for diarrheal illness affecting girls. For poor children and girls, this may reflect reduced household investment in care, which may result in increased risks of mortality.

As a leading cause of global child mortality, the primary impact of diarrheal disease is the health burden borne by children <5 years of age and their families [1]. However, there is also a growing awareness of the economic burden created by diarrheal disease. Several studies have attempted to estimate the economic burden of illness, especially that portion associated with the healthcare system [2–6]. Estimating these costs is critical for evaluating potential interventions to reduce the health burden, whether through vaccination, improved water and sanitation, or others [7], given that these costs can partially offset the required investment. Less is understood about the household economic burden associated with diarrhea. Although this may be small in absolute terms, it may be substantial relative to incomes of low-income households, resulting in reduced care seeking and worsening impoverishment.

Although most episodes of diarrheal illness can be treated inexpensively with timely diagnosis [8], evidence suggests that many low-income families lack access to high-quality, low-cost treatments for diarrheal illnesses or simply fail to utilize appropriate care [9, 10]. Reasons for such access and utilization barriers range from a lack of healthcare provision to poor transportation and even climate conditions [11, 12]. There is also reason to believe that access to and utilization of care for diarrheal illness may be driven by household economic and cost constraints [11, 13, 14, 15]. Conversely, reducing out-of-pocket expenses stimulates greater demand and utilization for healthcare [16, 17].

[a]Present affiliation: GlaxoSmithKline Biologicals, Global Medical Affairs, Wavre, Belgium.

Correspondence: Richard Rheingans, PhD, Department of Environmental and Global Health, University of Florida, Box 100188, 101 S Newell Dr, Gainesville, FL 32610 (rrheing@ufl.edu).

Clinical Infectious Diseases 2012;55(S4):S317–26

DOI: 10.1093/cid/cis763

Understanding these relationships is crucial for policymakers, particularly given health financing debates over user fees and other cost-sharing mechanisms [18]. For instance, although user fees may be effective options for discretionary care, they can have adverse impacts when applied to primary care or preventive services like diarrheal illness care [19]. Studies in several African settings have shown that reductions in user fees are more likely to stimulate demand for public healthcare services and that revenue collected from user fees are often not efficiently spent [17, 19–21]. Mwabu et al [19] found that during a period of modest cost sharing in Kenyan public clinics, demand dropped by nearly 50 percent.

This paper explores these issues using baseline survey data collected from 3 of the 4 African sites (Kenya, Gambia, and Mali) participating in the Global Enteric Multicenter Study (GEMS) prior to the onset of the main GEMS case/control study. The health economics substudy has 3 related objectives: to (1) estimate and characterize household costs associated with childhood diarrhea episodes by type and setting; (2) explore how child and household characteristics alter cost patterns; and (3) explore whether and how high costs can serve as a barrier to care or contribute to impoverishment of the household.

METHODS

This study uses data from the GEMS on acute diarrheal care in 3 African countries—Kenya, Gambia, and Mali. These countries were chosen in part owing to their relatively high rates of diarrheal illnesses and early childhood mortality. The sampling of households as part of the baseline Health Utilization and Attitudes Survey (HUAS) that preceded onset of the GEMS case/control study is described by Kotloff et al in this supplement [22]. Retrospective data were collected on household costs for children <5 years of age with diarrhea in the previous 2 weeks. Data were collected using a standardized interview from an age-stratified random sample of approximately 1000 households containing a child 0–59 months of age within each study area (described in [23]). Analyses were weighted on the basis of probability of selection. Each site aimed to enroll 400 infants 0–11 months of age, 370 children 12–23 months of age, and 370 children 24–59 months of age.

Sample sizes varied among countries and are presented in Table 1. The initial household sampling was expected to be large enough to identify approximately 200 children with diarrhea during the previous 2 weeks and 150 children with

Table 1. Study Population Characteristics and Subsamples

	Gambia						Kenya						Mali					
	Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs		Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs		Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs	
Sex																		
Male	149	(57)	56	(62)	49	(62)	157	(57)	97	(55)	78	(57)	68	(54)	46	(54)	44	(52)
Female	111	(43)	35	(39)	30	(38)	119	(43)	78	(45)	60	(44)	58	(46)	40	(47)	40	(48)
Maternal Education																		
None to primary	48	(19)	15	(17)	12	(15)	145	(53)	93	(53)	75	(54)	74	(59)	51	(59)	51	(61)
Finished primary	7	(3)	5	(6)	5	(6)	120	(44)	74	(42)	56	(41)	11	(9)	7	(8)	6	(7)
Some secondary	2	(1)	1	(1)	1	(1)	11	(4)	8	(5)	7	(5)	9	(7)	6	(7)	6	(7)
Religious only	203	(78)	70	(77)	61	(77)	…	…	…	…	…	…	32	(25)	22	(26)	21	(25)
Age																		
0–11 mo	96	(37)	34	(37)	32	(41)	116	(42)	66	(38)	52	(38)	44	(35)	29	(34)	28	(33)
12–23 mo	99	(38)	30	(33)	25	(32)	103	(37)	69	(39)	55	(40)	56	(44)	38	(44)	38	(45)
24–59 mo	65	(25)	27	(30)	22	(28)	57	(21)	40	(23)	31	(23)	26	(21)	19	(22)	18	(21)
Severity																		
Mild	47	(18)	18	(20)	16	(20)	93	(34)	44	(25)	30	(22)	47	(37)	26	(30)	26	(31)
Moderate/severe	213	(82)	73	(80)	63	(80)	183	(66)	131	(75)	108	(78)	79	(63)	60	(70)	58	(69)
Duration																		
1–3 d	75	(43)	30	(44)	25	(42)	85	(40)	59	(39)	52	(43)	50	(54)	33	(54)	33	(56)
4–7 d	93	(53)	35	(52)	32	(54)	115	(54)	80	(53)	61	(50)	40	(43)	26	(43)	24	(41)
8–14 d	6	(3)	3	(4)	2	(3)	13	(6)	11	(7)	8	(7)	2	(2)	1	(2)	1	(2)
15+ d	…	…	…	…	…	…	1	(1)	0	(0)	0	(0)	1	(1)	1	(2)	1	(2)

All data are presented as No. (%).

household costs associated with the episodes. Based on World Health Organization methods for estimating diarrheal costs [8], this was expected to be sufficient to produce estimates of means with a confidence interval of ±10% of the true mean with 80% power. The observed power in each country varies based on the variance in costs within each and the actual number of episodes. Sample sizes were not powered for secondary analyses to detect differences among subgroups.

We examined direct medical, direct nonmedical, and indirect costs. Direct medical costs (eg, medications, visits, diagnostics) were defined as either informal or formal expenditures, with the former representing care provided by a local healer or provider and the latter combining both outpatient and inpatient care. Outpatient and inpatient facilities at each site are described in more detail in Kotloff et al [22], but outpatient facilities were primarily health centers and private doctors' offices, while inpatient facilities were primarily public district hospitals. Direct nonmedical costs were broken down by transportation and other costs, whereas indirect costs were based on time lost from income-generating employment. For both medical and total costs, some cases incur no costs and the remainder of episodes typically produce a right-skewed distribution. Descriptive statistics (means and standard errors) for costs are provided for all cases, those incurring medical or other costs, and the proportion incurring costs (Table 1). Costs were collected in local currencies, converted to US$, and adjusted to 2011 as the reference year.

We also examined how child, household, and episode characteristics were associated with the costs incurred by households. This was analyzed separately for direct medical costs and total costs. Analysis of variance was used to assess the effect of household economic status, maternal education, child sex, age, duration of illness, and illness severity. Multivariate analysis was considered, but not presented owing to the limited sample size. This analysis was conducted separately for all episodes and those incurring medical or any costs. Logistic regression was used to estimate the effect of these variables on the likelihood of costs being incurred by the household. Household economic status is based on an asset index calculated using principal components analysis using the full household sample for each country [24]. Maternal education was broken down into 4 categories: none to some primary, completed primary education, some secondary education, and religious education only. Because of the limited sample size and power, we considered $P < .05$ as statistically significant and P values between .05 and .20 as marginally significant.

Given the empirical evidence citing costs as a significant factor driving healthcare behavior and utilization, we examined the potential impact of costs on household impoverishment and avoidance of care due to economic costs. This is done by examining respondents' self-reported reasons for not seeking care and strategies for paying for the costs. We also examined the distribution of costs to households and the possibility of large expenditures.

RESULTS

Expenditures by Type and Category

Table 2 displays costs of diarrheal episodes by type (direct medical, direct nonmedical, and indirect), type of medical (consultations, medications, and diagnostic), and setting of care (formal and informal). Mean costs and standard errors are calculated for all episodes and for those incurring a cost and listed for each category.

Of respondents reporting an episode of diarrhea in the previous 2 weeks, 35%, 65%, and 68% incurred some costs in the GEMS sites in Gambia, Kenya, and Mali, respectively. The mean total household costs per episode ranged from $2.63 in Gambia to $6.24 in Kenya, and the total cost among those with nonzero costs ranged from $6.01 in Mali to $8.83 in Kenya. Direct medical costs accounted for 11%, 27%, and 54% of that total cost in Kenya, Gambia, and Mali, respectively. Household indirect costs (productivity losses) accounted for more than half of the total cost in Gambia and Kenya and somewhat less (42%) in Mali. In Gambia and Kenya, expenditure on care from informal providers was more than that of formal providers. In Mali, expenditure on informal care was even greater than in Gambia or Kenya, but only accounted for 24% of the direct medical expenditure. In all 3 countries, medications (whether medically indicated or not) accounted for the majority of the direct medical cost, ranging from 77% in Gambia to 86% in Kenya.

In addition to mean costs, we examined the distribution of costs to better understand how high-cost events might affect households. The distributions of total costs by wealth quintile for each country are shown in Figure 1. In Gambia, among all children 25% of episodes resulted in costs over $1.73, in 10% the cost was over $6.50, and in 5% it was over $15.27. In Kenya the distribution was higher with 25% having costs over $4.93, 10% having costs over $11.05, and 5% having costs over $21.20. In Mali the costs were similar, with 25% over $4.26, 10% over $13.84, and 5% over $20.77.

Determinants of Costs

We examined the effect of household economic status, maternal education, child sex, child age, disease severity, and disease duration on the likelihood of incurring direct medical costs and the mean household cost (for all episodes and those incurring costs) for each of the 3 countries (Table 3). For each determinant the table shows the probability of incurring a cost and the mean household cost. P values represent a bivariate comparison of differences among the different subgroups for

Table 2. Household Costs Associated With Diarrheal Illness by Type and Setting (2011 US$) in Gambia, Kenya, and Mali

	Gambia				Kenya				Mali			
	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error
Cost by Type	n = 259		n = 97		n = 275		n = 186		n = 126		n = 86	
Direct medical	0.71	0.16	1.81	0.39	0.70	0.09	0.99	0.13	2.20	0.44	3.22	0.61
Direct nonmedical	0.37	0.07	0.96	0.16	0.55	0.28	0.79	0.39	0.19	0.05	0.28	0.07
Total direct	1.08	0.18	2.76	0.40	1.25	0.30	1.77	0.41	2.39	0.46	3.50	0.65
Indirect cost	1.55	0.42	3.97	1.03	4.99	1.41	7.06	1.97	1.72	0.40	2.52	0.56
Total	2.63	0.53	6.74	1.24	6.24	1.45	8.83	2.01	4.11	0.67	6.01	0.91
Direct medical cost by setting												
Informal (healer, pharmacist)	0.49	0.15	1.25	0.37	0.41	0.06	0.59	0.09	0.60	0.12	0.87	0.17
Formal (hospital, clinic, office, etc)	0.22	0.06	0.56	0.16	0.28	0.06	0.40	0.09	1.60	0.39	2.34	0.56
Direct medical cost by purpose												
Consultation	0.07	0.04	0.17	0.10	0.05	0.02	0.08	0.02	0.29	0.06	0.41	0.08
Medication	0.55	0.11	1.40	0.27	0.60	0.08	0.85	0.10	1.81	0.29	2.66	0.35
Diagnostic tests	0.09	0.09	0.24	0.23	0.05	0.02	0.07	0.03	0.10	0.07	0.15	0.11

each determinant. For many of the determinants, low sample size in the subcategories (Table 1) led to limited statistical power.

In all 3 countries, there was a trend toward differences by economic status for both medical (Table 3) and total costs (Table 4). For medical costs, there were increased expenditures for children in high-wealth quintiles. For Gambia and Kenya this was found for both mean costs among all episodes and among episodes with nonzero costs. For Mali, direct medical costs exhibit an inverted U-shaped curve with lowest costs among the poorest quintile and highest costs among children in the middle and upper wealth quintiles for both all episodes and those with nonzero costs. These trends were statistically significant ($P < .05$) or marginally significant ($P < .20$). In all 3 countries, medical costs per episode were 2–3 times greater in the highest wealth quintile compared with the lowest. Trends were similar for total household, but only statistically significant for Mali and Gambia.

In Mali and Gambia there were significant or marginally significant differences in household medical and total costs by sex. For both countries, household direct medical and total costs for boys were approximately twice that for girls; however, the differences were only marginally statistically significant. For Kenya there were no differences by sex.

Although there were country-level differences in medical and total costs by maternal education, there were few clear patterns within or among countries. There were no clear associations between child age and household medical or total costs within or across countries.

In Kenya and Mali, there were higher household medical costs for moderate-to-severe episodes (all episodes and those with nonzero costs). However, there was no such association for Gambia. Total household costs were higher for moderate-to-severe cases only in Mali. Duration of illness was also associated with household medical and total costs in Gambia and Mali.

Costs as a Barrier and Cause of Impoverishment

Table 5 shows household reasons for not seeking care or hospitalization and payment method for the expenses associated with the episode. Across all 3 countries the main reasons for not seeking hospital care when advised were either not believing their child needed care or that the costs were too high. Broken down further, 55.8% of Kenyan households seeking alternative forms of care did so because they felt hospital treatment or transportation costs were too high; nearly 18% did not think their child was sick enough to seek hospital care. Among Gambian households, these figures are 22.2% and 48.1%, respectively. Among Malian households, they are roughly 53% and 30%.

Similar results were found among those not seeking any care. Among all 3 countries, the most common reasons for not seeking any care was that, on average, 53.4% of all

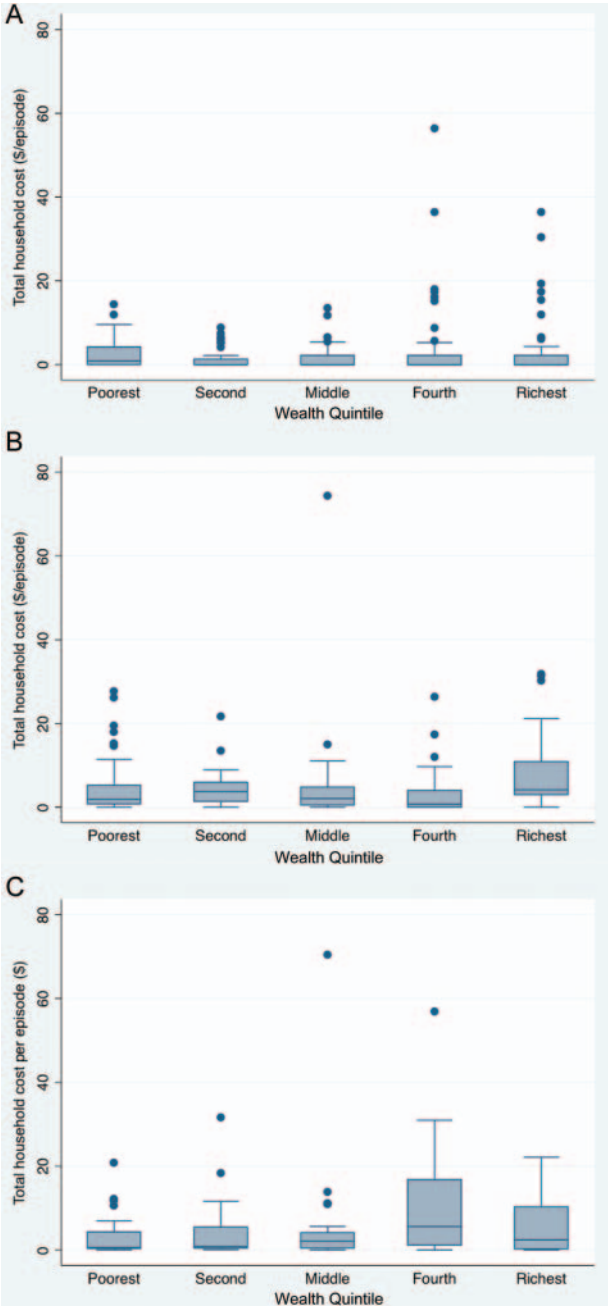

Figure 1. Distribution of total household diarrhea costs by wealth quintile ($/episode) in Gambia (*A*), Kenya (*B*), and Mali (*C*). For each group the box shows the 25th and 75th percentile and the bars show the 5th and 95th percentiles.

households believed their child did not need care for his or her illness. Among Kenyan families, treatment and transportation costs were close behind (41.2%), followed by a high demand for traditional medicine (17.4%), too far a distance (12.7%), and lack of transportation (9.5%). For Gambia, these included treatment costs (22.5%), transportation costs (10%), and preference for traditional medicine (10%). For Malian households, treatment and transportation costs (26.6%) and preference for traditional medicine (10%) were also common reasons. The data indicate that households either believe their child does not need care or, if he or she does, costs are too high.

DISCUSSION

The GEMS case/control study, the keystone of GEMS, is intended to provide information on the etiology and burden of moderate-to-severe diarrhea and its nutritional and mortality consequences. However, as part of the rationale for undertaking GEMS, we also wished to expand the assessment of burden by gathering information on the direct and indirect economic costs of diarrheal disease in sites where the case/control study would be carried out. Our results document a substantial economic burden stemming from diarrheal disease and provide an additional reason to support interventions to control the incidence and severity of diarrheal disease.

What Are the Costs and Where Do They Occur?

Our results suggest that households encounter a substantial economic burden due to childhood diarrhea in the 3 settings. For episodes with nonzero costs, the mean total cost ranged from $6.01 in Mali to $8.83 in Kenya. When all episodes are considered, the range was $2.63 in Gambia to $6.24 in Kenya. Although these amounts may seem small in absolute terms, these are settings where a substantial portion of households live on <$1 per day. In addition, diarrhea is frequent in children <5 years of age [25], implying that these expenses may be incurred regularly.

Direct medical expenses only account for a fraction of these total costs: 27% in Gambia, 11% in Kenya, and 53% in Mali. Costs in informal settings ranged from $0.41 in Kenya to $0.60 in Mali per episode, and accounted for more than half of the household medical costs in both Gambia and Kenya. In Mali, direct medical costs in formal settings accounted for a larger fraction of household costs. For all 3 countries, the majority of household direct medical costs were for medications. High nonmedical costs, whether for transportation or for lost earnings, suggest that user fees for formal care may not be the only financial barriers to treatment.

What Are the Determinants for Household Costs?

While the patterns vary among countries, wealth and sex appear to be associated with direct medical and total household diarrheal costs. Although there are a number of potential explanations for this association, the relationship between household wealth and diarrhea economic burden may reflect rationing of care in poorer households. That is, household resources provide

Table 3. Household Direct Medical Costs for Childhood Diarrhea by Socioeconomic, Demographic, and Illness Characteristic in Gambia, Kenya, and Mali

	Gambia						Kenya						Mali					
	Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs	
	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA
Wealth quintile																		
Poorest	0.40		0.40		1.33		0.50		0.44		1.22		1.11		0.72		1.55	
Second	0.29	*P* = .07	0.21	.07	1.37	*P* = .04	0.52	*P* = .03	0.59	.22	0.92	*P* = .09	2.15	*P* = .38	0.75	.79	2.85	*P* = .26
Middle	1.24		0.35	.68	3.91		0.60		0.47	.79	1.20		3.16		0.68	.76	4.65	
Fourth	0.42		0.31	.40	2.01		0.63		0.54	.36	1.19		2.73		0.47	.09	5.84	
Richest	1.55		0.30	.37	7.76		1.42		0.66	.06	2.44		2.01		0.65	.61	3.53	
Sex																		
Male	1.00		0.34		4.13		0.67		0.50		1.42		2.75		0.65		4.37	
Female	0.37	*P* = .10	0.29	.48	1.49	*P* = .10	0.74	*P* = .62	0.53	.67	1.37	*P* = .87	1.52	*P* = .15	0.68	.73	2.26	*P* = .10
Maternal education																		
None to primary	0.41		0.19		3.47		0.68		0.57		1.18		2.76		0.70		4.03	
Finished primary	1.48	*P* = .99	0.91	.00	3.57	*P* = .92	0.74	*P* = .78	0.46	.14	1.74	*P* = .33	0.26	*P* = .57	0.55	.33	0.47	*P* = .66
Some secondary	0.22		0.50	.33	0.43		0.53		0.45	.48	1.56		1.37		0.60	.58	2.56	
Religious only	0.75		0.31	.14	2.79								1.92		0.63	.47	3.07	
Age																		
0–11 mo	0.91		0.33		3.42		0.48		0.45		1.11		1.86		0.64		3.03	
12–23 mo	1.01	*P* = .56	0.26	.28	4.45	*P* = .40	0.86	*P* = .13	0.53	.22	1.68	*P* = .26	4.11	*P* = .12	0.68	.66	6.22	*P* = .10
24–59 mo	0.49		0.34	.89	1.98		0.71		0.55	.23	1.35		1.08		0.69	.64	1.56	
Severity																		
Mild	0.81		0.35		2.47		0.22		0.34		0.59		0.71		0.58		1.24	
Moderate/severe	0.68	*P* = .66	0.31	.61	3.01	*P* = .65	0.92	*P* < .001	0.60	.00	1.65	*P* = .003	3.13	*P* = .02	0.72	.15	4.53	*P* = .06
Duration																		
1–3 d	0.51		0.30		2.19		0.71		0.64		1.14		2.31		0.70		3.37	
4–7 d	0.75	*P* = .10	0.39	.35	2.47	*P* = .10	0.84	*P* = .91	0.53	.20	1.63	*P* = .59	1.81	*P* = .06	0.53	.12	3.68	*P* = .14
8–14 d	4.28		0.57	.27	7.52		0.76		0.66	.86	1.29		3.84		0.29	.24	13.21	
15+ d	…		…		…		…		…		…		16.19		1.00		16.19	

Abbreviation: ANOVA, analysis of variance.

Table 4. Household Total Costs for Childhood Diarrhea by Socioeconomic, Demographic, and Illness Characteristic in Gambia, Kenya, and Mali

	Gambia						Kenya						Mali					
	Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs	
	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA
Wealth quintile																		
Poorest	2.65		0.52		5.04		5.39		0.67		5.93		3.20		0.72		4.45	
Second	1.04	*P* = .09	0.23	.01	3.33	*P* = .03	6.16	*P* = .89	0.76	.40	4.83	*P* = .12	3.50	*P* = .52	0.75	.79	4.64	*P* = .28
Middle	1.38		0.36	.21	3.81		5.31		0.59	.40	7.08		4.00		0.68	.76	5.89	
Fourth	2.84		0.40	.29	6.96		8.99		0.59	.42	4.10		5.89		0.54	.21	10.98	
Richest	5.28		0.31	.08	16.96		6.09		0.78	.26	7.83		4.54		0.70	.92	6.45	
Sex																		
Male	3.41		0.39		8.69		7.93		0.62		7.19		5.20		0.69		7.57	
Female	1.72	*P* = .16	0.35	.65	4.52	*P* = .25	4.11	*P* = .40	0.71	.18	4.87	*P* = .29	2.77	*P* = .09	0.68	.93	4.07	*P* = .07
Maternal education																		
None to primary	1.06		0.26		4.12		4.46		0.69		6.41		4.38		0.70		6.22	
Finished primary	10.45	*P* = .36	0.91	.00	11.49	*P* = .82	7.48	*P* = .04	0.63	.43	5.25	*P* = .03	0.79	*P* = .52	0.65	.73	1.21	*P* = .52
Some secondary	0.22		0.50	.47	0.43		14.27		0.62	.71	10.08		2.42		0.60	.58	4.02	
Religious only	2.66		0.38	.19	6.84								5.33		0.67	.74	7.96	
Age																		
0–11 mo	1.87		0.35		5.39		4.35		0.57		5.65		4.02		0.66		6.10	
12–23 mo	3.73	*P* = .36	0.31	.55	11.33	*P* = .17	4.71	*P* = .21	0.67	.13	5.49	*P* = .70	6.23	*P* = .23	0.68	.84	9.17	*P* = .17
24–59 mo	2.33		0.42	.39	5.61		8.32		0.71	.08	6.67		2.39		0.73	.53	3.27	
Severity																		
Mild	2.84		0.42		6.82		4.92		0.47		4.01		1.82		0.58		3.15	
Moderate/severe	2.58	*P* = .93	0.36	.59	6.86	*P* = .97	6.85	*P* = .98	0.75	<.01	6.69	*P* = .38	5.55	*P* = .02	0.75	.06	7.39	*P* = .09
Duration																		
1–3 d	2.48		0.41		5.62		5.44		0.70		6.61		3.75		0.70		5.38	
4–7 d	2.47	*P* < .001	0.43	.82	5.72	*P* < .001	8.75	*P* = .93	0.73	.71	5.80	*P* = .80	4.79	*P* = .38	0.59	.34	8.10	*P* = .43
8–14 d	14.48		0.70	.18	20.57		4.88		0.87	.22	5.63		5.39		0.29	.24	18.53	
15+ d	…		…		…		…		…		…		18.32		1.00		18.32	

Abbreviation: ANOVA, analysis of variance.

Table 5. Reasons for Not Seeking Treatment and Sources of Household Costs for Diarrhea Episodes—Kenya, Gambia, and Mali

Question	Kenya (n = 63)	Gambia (n = 49)	Mali (n = 30)
Why did households not seek care for their child?			
No need for care	44.4%	49.0%	66.7%
Distance too far	12.7%	2.0%	0.0%
Lack of transportation	9.5%	2.0%	0.0%
No time off work	3.2%	8.2%	0.0%
Local situation (political)	0.0%	4.1%	3.3%
Transportation costs	4.8%	10.2%	3.3%
Treatment costs	36.5%	22.5%	23.3%
Leaving other children at home	4.8%	0.0%	0.0%
Unhappy with clinical services	1.6%	8.2%	0.0%
Preferred traditional medicine	17.4%	10.2%	10.0%
Cultural differences	3.1%	0.0%	0.0%
Other	12.6%	0.0%	0.0%
Why did the household not seek hospital care when advised?	Kenya (n = 34)	Gambia (n = 27)	Mali (n = 17)
Hospital too far	0.0%	7.4%	0.0%
No transportation	2.9%	3.7%	0.0%
Travel costs too high	2.9%	14.8%	0.0%
No time off work	0.0%	11.1%	0.0%
Local situation	0.0%	0.0%	0.0%
Treatment costs	52.9%	7.4%	52.9%
Needs of other children at home	2.9%	0.0%	0.0%
Child not sick enough	17.7%	48.2%	29.4%
Unhappy with clinical services	0.0%	3.7%	0.0%
Other	20.6%	3.7%	17.7%
Where did the money come from?	Kenya (n = 213)	Gambia (n = 211)	Mali (n = 96)
Fewer meals	18.3%	7.1%	8.3%
Cutting other expenses	15.0%	17.1%	12.5%
Savings	34.3%	44.1%	65.6%
Borrowing	15.5%	8.1%	3.1%
Selling assets	16.4%	1.9%	5.2%
Donations	1.4%	0.0%	3.1%
Relative or friend	9.4%	5.7%	3.1%
Other	7.9%	6.2%	3.1%

a constraint on what can be spent on treatment or transportation, resulting in less care seeking and less expenditure among poor households. However, we saw no differences in the proportion of episodes incurring some costs among wealth quintiles, suggesting that household wealth may not affect whether money is spent, but rather how much is spent.

Sex was a second determinant of household diarrhea costs in Gambia and Mali, but not in Kenya. There are 2 potential explanations for this association. First, it is possible that this reflects differences in diarrhea severity between boys and girls that result in the need for greater care among boys. However, there were no differences in the frequency of moderate-to-severe diarrhea between boys and girls in either country. The second interpretation is that cost differences reflect intrahousehold resource allocation that disadvantages girls. Several studies have documented reduced health expenditures for girls in low-income settings [26–29]. If resources are limiting care seeking for diarrhea, then it is plausible that girls will bear a greater burden in terms of missed treatment and the resulting negative outcomes. It is interesting to note that this relationship only held for the 2 lower-income countries in the study. It is also interesting to note that this relationship was not observed in our related study in 3 Asian settings [30].

Is Household Economic Burden a Barrier to Appropriate Care?

Average costs per episode only provide one aspect of the burden costs place on households in low-income settings. Three other related factors must be considered: the distribution of costs, the potential for impoverishment due to the costs, and the health burden of avoided costs. The cost

distributions within each setting demonstrate that costs often substantially exceed the mean. In all countries, 10% of episodes resulted in costs that were twice the mean and even further above the median. In Kenya and Mali, this resulted in 10% of cases having costs of >$10, a substantial burden in settings where households live on $1 per day. Figure 1 shows that even the poorest households experience episodes with high costs. Additionally, diarrhea is a frequent outcome among children <5 years of age, implying that each episode brings a chance of high costs when compared to earnings. In all 3 countries, households were most likely to get the funds from reduced savings. Other common responses included incurring debt and selling household assets. Our results do not allow us to determine the long-term consequences of these costs on household impoverishment. However, it is likely that reduced savings, diminished assets, and increased debt would make it harder for households to respond to adverse economic events in the future, especially for the small but important fraction of households that incur substantial costs.

Possibly the greatest economic burden is not the costs themselves, but that they may encourage rationing of care for children with diarrhea. The most common reasons for not seeking care was related to a lack of resources or a perception that the episode was not severe. These costs were not just the formal costs of direct medical treatment but also the costs of transportation, childcare, and missed work. Given that direct medical costs in formal settings account for only a small fraction of household costs, it is unclear whether reduced user fees would have an impact on this barrier. Medication costs (typically separate from user fees) are substantial, suggesting that even with low costs for visits, households face other economic costs that may impede access. Lower observed costs for girls and children in poor households are likely symptoms of this rationing of care, implying that the health burden associated with household economic costs falls primarily on these children. The data analyzed here do not allow us to directly address whether these household costs resulted in greater adverse outcomes (eg, severe illness or mortality); however, the results point to the importance of addressing these questions empirical with the additional data being collected in the study.

Limitations

The current work suffers from several limitations. First, the study sample size was designed to provide estimates of overall costs within a margin of error but was not powered to examine determinants of costs. As a result, differences among subgroups are often not statistically significant and could be addressed with larger samples in subsequent research. Second, one-time cross-sectional data did not allow directly examination of the long-term consequences of incurred costs by household for individual events or repeated episodes. Last, the cross-sectional nature of the study makes it difficult to assess whether low costs for specific subgroups are the result of reduced severity, cheaper services, or rationing of care. Additional work must also be conducted to better understand how the complex interaction between direct medical, direct nonmedical, and indirect costs impact households' demand for and decisions to seek informal or formal care.

CONCLUSIONS

Diarrheal episodes are common among children <5 years of age in low-income settings, resulting in significant mortality burden as well as substantial economic costs associated with nonfatal events. These 2 aspects of burden—mortality and household costs—may be closely connected. Costs may serve as barriers that result in reduced healthcare seeking, especially for poorer households and for girls. These costs may force households to take other steps like borrowing and reducing savings that may expose them to economic insecurity. While the results here cannot prove this connection between household costs and mortality, it points to importance of further study. The costs of diarrhea treatment to the healthcare system are important and must be considered by national decision makers choosing between health interventions.

Notes

Acknowledgments. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874). We wish to thank Ciara O'Reilly and Eric Mintz of the US Centers for Disease Control and Prevention for their contributions to study design and data collection.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the US Centers for Disease Control and Prevention.

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet **2010**; 375:1969–87.
2. Atherly D, Dreibelbis R, Parashar UD, Levin C, Wecker J, Rheingans RD. Rotavirus vaccination: cost-effectiveness and impact on child mortality in developing countries. J Infect Dis **2009**; 200(suppl 1):S28–38.
3. Khoury H, Ogilvie I, El Khoury AC, Duan Y, Goetghebeur MM. Burden of rotavirus gastroenteritis in the Middle Eastern and North African pediatric population. BMC Infect Dis **2011**; 11:9.
4. Kirigia JM, Sambo LG, Yokouide A, Soumbey-Alley E, Muthuri LK, Kirigia DG. Economic burden of cholera in the WHO African region. BMC Int Health Hum Rights **2009**; 9:14.

5. Rheingans R, Atherly D, Anderson J. Distributional impact of rotavirus vaccination in 25 GAVI countries: estimating disparities in benefits and cost-effectiveness. Vaccine **2012**; 30(suppl 1):A15–23.
6. Tate JE, Rheingans RD, O'Reilly CE, et al. Rotavirus disease burden and impact and cost-effectiveness of a rotavirus vaccination program in Kenya. J Infect Dis **2009**; 200(suppl 1):S76–84.
7. Rheingans RD, Antil L, Dreibelbis R, Podewils LJ, Bresee JS, Parashar UD. Economic costs of rotavirus gastroenteritis and cost-effectiveness of vaccination in developing countries. J Infect Dis **2009**; 200(suppl 1): S16–27.
8. World Health Organization. Guidelines for estimating the economic burden of diarrheal disease with focus on assessing the costs of rotavirus diarrhea. Geneva, Switzerland: WHO, **2005**.
9. Patil AV, Somasundaram KV, Goyal RC. Current health scenario in rural India. Aust J Rural Health **2002**; 10:129–35.
10. Rao NS. Delivery of health services in the public sector. Financing and delivery of health care services. In: India in National Commission on Macroeconomics and Health - Background papers. New Delhi: Ministry of Health and Family Welfare, **2005**; 4:43–59.
11. Gelberg L, Andersen RM, Leake BD. The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. Health Serv Res **2000**; 34:1273–302.
12. Banerjee A, Deaton A, Duflo E. Health, health care, and economic development: Wealth, health, and health services in rural Rajasthan. Am Econ Rev **2004**; 94:326–30.
13. Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski GE, eds. Changing the US health care systems: key issues in health services policy and management. San Francisco: Jossey-Bass, **2001**:3–30.
14. Chola L, Robberstad B. Estimating average inpatient and outpatient costs and childhood pneumonia and diarrhoea treatment costs in an urban health centre in Zambia. Cost Eff Resour Alloc **2009**; 7:16.
15. Burton DC, Flannery B, Onyango B, Larson C, Alaii J, Zhang X, Hamel MJ, Breiman RF, Feikin DR. Healthcare-seeking behaviour for common infectious disease-related illnesses in rural Kenya: a community-based house-to-house survey. J Health Popul Nutr **2011**; 29:61–70.
16. Ridde V, Morestin F. A scoping review of the literature on the abolition of user fees in health care services in Africa. Health Policy Plan **2011**; 26:1–11.
17. James CD, Hanson K, McPake B, et al. To retain or remove user fees? Reflections on the current debate in low- and middle-income countries. Appl Health Econ Health Policy **2006**; 5:137–53.
18. Stanton B, Clemens J. User fees for health care in developing countries: a case study of Bangladesh. Soc Sci Med **1989**; 29:1199–205.
19. Mwabu G, Mwanzia J, Liambila W. User charges in government health facilities in Kenya: effect on attendance and revenue. Health Policy Plan **1995**; 10:164–70.
20. Chuma J, Musimbi J, Okungu V, Goodman C, Molyneux C. Reducing user fees for primary health care in Kenya: policy on paper or policy in practice? Int J Equity Health **2009**; 8:15.
21. Gilson L. The lessons of user fee experience in Africa. Health Policy Plan **1997**; 12:273–85.
22. Kotloff KL BW, Nasrin D, et al. The Global Enterics Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis **2012**; 55(Suppl 4):S232–45.
23. Levine MM. Etiology of diarrhoeal and enteric diseases in developing countries: the Global Enteric Multicenter Study (GEMS) project. Global vaccine research forum. Bamako, Mali: World Health Organization, **2009**.
24. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. Measurement of living standards. In: O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M, eds. Analyzing health equity using household survey data: a guide to techniques and their implementation. Washington, DC: World Bank, **2008**:69–82.
25. Feikin DR, Olack B, Bigogo GM, et al. The burden of common infectious disease syndromes at the clinic and household level from population-based surveillance in rural and urban Kenya. PLoS One **2011**; 6:e16085.
26. Pokhrel S, Snow R, Dong H, Hidayat B, Flessa S, Sauerborn R. Gender role and child health care utilization in Nepal. Health Policy **2005**; 74:100–9.
27. Pandey A, Sengupta PG, Mondal SK, et al. Gender differences in healthcare-seeking during common illnesses in a rural community of West Bengal, India. J Health Popul Nutr **2002**; 20:306–11.
28. Nuruddin R, Hadden WC, Petersen MR, Lim MK. Does child gender determine household decision for health care in rural Thatta, Pakistan? J Public Health (Oxf) **2009**; 31:389–97.
29. Nguyen H, Knowles J. Demand for voluntary health insurance in developing countries: the case of Vietnam's school-age children and adolescent student health insurance program. Soc Sci Med **2010**; 71:2074–82.
30. Rheingans R, Kukla M, Faruque A, et al. Determinants of household costs associated with childhood diarrhea in three South Asian settings. Clin Infect Dis **2012**; 55(Suppl 4):S327–35.

SUPPLEMENT ARTICLE

Determinants of Household Costs Associated With Childhood Diarrhea in 3 South Asian Settings

Richard Rheingans,[1] Matt Kukla,[1] Abu Syed Golam Faruque,[2] Dipika Sur,[3] Anita K. M. Zaidi,[4] Dilruba Nasrin,[5] Tamer H. Farag,[5] Myron M. Levine,[5] and Karen L. Kotloff[5]

[1]Department of Environmental and Global Health, University of Florida, Gainesville; [2]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; [3]National Institute of Cholera and Enteric Diseases, Kolkata, India; [4]Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan; and [5]Center for Vaccine Development, University of Maryland School of Medicine, Baltimore

In addition to being a major cause of mortality in South Asia, childhood diarrhea creates economic burden for affected households. We used survey data from sites in Bangladesh, India, and Pakistan to estimate the costs borne by households due to childhood diarrhea, including direct medical costs, direct nonmedical costs, and productivity losses. Mean cost per episode was $1.82 in Bangladesh, $3.33 in India, and $6.47 in Pakistan. The majority of costs for households were associated with direct medical costs from treatment. Mean costs understate the distribution of costs, with 10% of cases exceeding $6.61, $8.07, and $10.11 in Bangladesh, India, and Pakistan, respectively. In all countries there was a trend toward lower costs among poorer households and in India and Pakistan there were lower costs for episodes among girls. For both poor children and girls this may reflect rationing of care, which may result in increased risks of mortality.

Diarrhea is a leading cause of child mortality in south Asia and globally, especially in low-income settings [1]. In addition to the health burden from mortality, diarrhea can have an important economic impact on the households of affected children and society as a whole. The economic costs to the healthcare system and governments can help offset the costs of interventions to reduce diarrheal morbidity and mortality, and there have been growing efforts to estimate these costs [2–8]. In addition, households themselves can bear a substantial economic burden due to the costs of treatment, other out-of-pocket expenses like transportation, and lost time from work. However, there is little empirical evidence of the magnitude of these costs to households.

Correspondence: Richard Rheingans, Department of Environmental and Global Health, University of Florida, Box 100188, 101 S Newell Dr, Gainesville, FL 32610 (rrheing@ufl.edu).

Clinical Infectious Diseases 2012;55(S4):S327–35

DOI: 10.1093/cid/cis764

Household economic costs of diarrhea episodes can have both an economic and a health impact [9–11]. First, they can reduce household resources for other activities, including productive investments, education, and food. Second, they can influence whether households seek care and how much they seek; that is, households—especially poor households—may avoid treatment due to the high cost, potentially exposing their children to higher risks of mortality. Given the high frequency of diarrhea among children in low-income settings, households are repeatedly balancing these health risks and economic costs.

This balance between household costs and health risks also interacts with important policy debates on user fees for treatment. Globally there has been a push to increase user fees for basic health services as a way to increase the financial sustainability of government-supported health systems. However, shifting costs of basic health services may result in households delaying or forgoing treatment [11, 12]. For example, Stanton and Clemens argued that instituting user fees in government health clinics in Bangladesh could have detrimental health impacts, via reductions in utilization, on

some of the most medically vulnerable groups that utilize this system, namely, women, children, and the poor [13]. For health conditions like childhood diarrhea for which there are effective, low-cost solutions [14, 15], this may have adverse consequences [16–19].

We explore these issues within the context of 3 South Asian nations: Bangladesh, India, and Pakistan. We had 3 related objectives: (1) to estimate and characterize household costs associated with childhood diarrhea episodes by type and setting, (2) to explore how child and household characteristics alter cost patterns, and (3) explore how high costs can serve as a barrier to care or contribute to impoverishment of the household.

METHODS

This study uses data from the Global Enteric Multicenter Study (GEMS) of acute moderate-and-severe diarrhea in 3 Asian countries—Bangladesh, India, and Pakistan. Specifically, data come from the Healthcare Utilization and Attitudes Survey (HUAS). All 3 countries were chosen because of their relatively high rates of child mortality. Data were collected using a standardized interview from an age-stratified random sample of approximately 1000 households containing a child 0–59 months of age within each study area (described in [20]). Analyses were weighted based on the probability of selection. Each site aimed to enroll 400 infants 0–11 months of age, 370 children 12–23 months of age, and 370 children 24–59 months of age. For children aged 0–59 months who experienced a diarrheal episode in the previous 14 days, additional retrospective data were collected on household expenditures for medical care, other direct costs such as transportation, and time lost from paid work (indirect costs).

Sample sizes varied among countries. The initial household sampling was expected to be large enough to identify approximately 200 children with diarrhea in the previous 14 days and 150 children with household costs associated with the episodes. Based on World Health Organization methods for estimating diarrheal costs [15], this was expected to be sufficient to produce estimates of means with a confidence interval of +10% of the true mean with 80% power. The observed power in each country varies based on the variance in costs within each and the actual number of episodes. Sample sizes were not powered for secondary analyses to detect differences between subgroups. Sample sizes for each of the countries and subpopulations are provided in Table 1.

Table 1. Study Population Characteristics and Subsamples

	Bangladesh						India						Pakistan					
	Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs		Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs		Children With Diarrhea		Children With Any Costs		Children With Direct Medical Costs	
Sex																		
Male	52	(55)	44	(54)	44	(54)	44	(48)	30	(48)	29	(49)	178	(51)	120	(53)	109	(52)
Female	43	(45)	37	(46)	37	(46)	48	(52)	32	(52)	30	(51)	171	(49)	107	(47)	100	(48)
Education																		
None to primary	31	(33)	25	(31)	25	(31)	36	(39)	22	(36)	20	(34)	211	(61)	139	(61)	128	(61)
Finished primary	50	(53)	43	(53)	43	(53)	41	(45)	32	(52)	31	(53)	34	(10)	26	(12)	26	(12)
Some secondary	14	(15)	13	(16)	13	(16)	14	(15)	7	(11)	7	(12)	20	(6)	12	(5)	12	(6)
Religious only	…	…	…	…	…	…	1	(1)	1	(2)	1	(2)	84	(24)	50	(22)	43	(21)
Age																		
0–11 mo	36	(38)	32	(40)	32	(40)	42	(46)	31	(50)	30	(51)	154	(44)	107	(47)	98	(47)
12–23 mo	39	(41)	32	(40)	32	(40)	26	(28)	17	(27)	17	(29)	117	(34)	71	(31)	65	(31)
24–59 mo	20	(21)	17	(21)	17	(21)	24	(26)	14	(23)	12	(20)	78	(22)	49	(22)	46	(22)
Severity																		
Mild	22	(23)	18	(22)	18	(22)	26	(28)	13	(21)	13	(22)	55	(16)	31	(14)	27	(13)
Moderate/severe	73	(77)	63	(78)	63	(78)	66	(72)	49	(79)	46	(78)	294	(84)	196	(86)	182	(87)
Duration																		
1–3 d	39	(53)	32	(51)	32	(51)	70	(83)	46	(79)	44	(80)	87	(43)	60	(40)	58	(41)
4–7 d	25	(34)	23	(37)	23	(37)	10	(12)	8	(14)	8	(15)	86	(43)	67	(45)	63	(45)
8–14 d	8	(11)	7	(11)	7	(11)	3	(4)	3	(5)	2	(4)	15	(8)	13	(9)	10	(7)
≥15 days	1	(1)	1	(2)	…	…	1	(1)	1	(2)	1	(2)	13	(7)	9	(6)	9	(6)

All data are presented as no. (%).

We examined direct medical costs, direct nonmedical costs, indirect medical costs, and total costs per child for utilizing healthcare services to treat a given case of diarrhea, all of which are converted to US dollars. Direct medical costs were defined as either informal or formal expenditures, with the former representing care provided by a local healer or pharmacists and the latter combining both outpatient and inpatient care. Outpatient and inpatient facilities at each site are described in more detail in Levine et al [20], but outpatient facilities were primarily health centers and private doctors' offices, while inpatient facilities were primarily public district hospitals. Direct nonmedical costs were broken down by transportation and other costs, whereas indirect medical costs were either time costs or other costs. For both medical and total costs, some cases incur no costs and the remainder of episodes typically produced a right-skewed distribution. Descriptive statistics (means and standard errors) for costs are provided for all cases, those incurring medical or other costs, and the proportion incurring costs (Table 1). All costs were collected in local currencies, converted to US dollars, and then adjusted to 2011 as the reference year.

We also examined how child, household, and episode characteristics were associated with the costs incurred by households. This was analyzed separately for direct medical costs and total costs. Analysis of variance was used to assess the effect of household economic status, maternal education, child sex, age, duration of illness, and illness severity. Multivariate analysis was considered but not presented due to the limited sample size. This analysis was conducted separately for all episodes and those incurring medical or any costs. Logistic regression was used estimate the effect of these variables on the likelihood of costs being incurred by the household. Household economic status is based on an asset index calculated using principal components analysis using the full household sample for each country [21]. Maternal education was broken down into 4 categories: no formal or some primary education, completed primary education, some secondary, and religious education only. Due to the limited sample size and power, we considered $P < .05$ as statistically significant and P values between .05 and .20 as marginally significant.

Given the empirical evidence citing costs as a significant factor driving healthcare behavior and utilization, we examine the potential impact of costs on household impoverishment and avoidance of care owing to the economic costs. This is done by examining respondents' self-reported reasons for not seeking care and strategies for paying for the costs. In addition, we examine the distribution of costs to households and the possibility of large expenditures.

RESULTS

Expenditures by Type and Category

Among all diarrhea episodes, household costs ranged from $1.82 in Bangladesh to $6.47 per episode in India (Table 2). Among cases with nonzero costs, it was slightly higher, ranging from $2.13 in Bangladesh to $6.83 in Pakistan. In all

Table 2. Household Costs Associated With Diarrheal Illness by Type and Setting (2011 US$) in Bangladesh, India, and Pakistan

	Bangladesh				India				Pakistan			
	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error	All Seeking Care	Std. Error	Incurring Any Treatment Cost	Std. Error
Cost by type	n = 95		n = 81		n = 92		n = 62		n = 349		n = 232	
Direct medical	0.94	0.16	1.09	0.18	2.08	0.39	3.31	0.57	2.30	0.53	3.51	0.80
Direct nonmedical	0.25	0.07	0.29	0.08	0.25	0.07	0.40	0.11	0.25	0.07	0.37	0.11
Total direct	1.19	0.20	1.39	0.23	2.33	0.44	3.71	0.64	2.54	0.54	3.89	0.81
Indirect cost	0.63	0.22	0.74	0.26	1.00	0.25	1.60	0.36	3.93	2.07	2.94	0.77
Total	1.82	0.34	2.13	0.39	3.33	0.60	5.31	0.84	6.47	2.16	6.83	1.15
Direct medical cost by setting												
Informal (healer, pharmacist)	0.50	0.08	0.58	0.09	0.28	0.12	0.44	0.18	0.58	0.13	0.89	0.20
Formal (hospital, clinic, office, etc)	0.44	0.15	0.51	0.17	1.80	0.39	2.87	0.60	1.72	0.51	2.63	0.78
Direct medical cost by purpose												
Consultation	0.05	0.01	0.05	0.02	0.55	0.11	0.87	0.17	0.55	0.18	0.85	0.27
Medication	0.76	0.10	0.89	0.11	1.56	0.30	2.47	0.43	1.75	0.50	2.68	0.76
Diagnostic tests	0.12	0.09	0.14	0.11	0.05	0.03	0.07	0.05	0.02	0.01	0.03	0.02

3 countries, direct medical costs exceeded direct nonmedical and indirect costs, accounting for 52% of costs per episode in Bangladesh, 51% in Pakistan, and 62% in India. Of these direct medical costs to households, in India and Pakistan, care in formal settings accounted for the majority (87% and 75%, respectively). In Bangladesh, formal costs were less than informal. In all 3 settings, the cost of medications far exceeded visits and diagnostic tests, accounting for 75% of the direct medical cost in India, 76% in Pakistan, and 81% in Bangladesh. Indirect costs from lost earnings amounted 30% of the household cost in India, 61% in Pakistan, and 35% in Bangladesh.

In addition to mean costs, we examined the range and distribution of costs within each setting. Figure 1 shows the distribution of costs by wealth quintile. For each group, the box represents the 25th and 75th percentile and the bars show the 5th and 95th percentile. In Bangladesh, 25% of episodes (or all cases and all wealth quintiles) had total household costs in excess of $1.94, 10% exceeded $6.61, and 5% were greater than $10.44. In India the range was higher: 25% greater than $5.88, 10% greater than $8.07, and 5% greater than $12.22. Household costs were similar in Pakistan with 25% greater than $4.15, 10% greater than $10.11, and 5% greater than $18.29.

Determinants of Costs

We examined the association between wealth, sex, education, age, severity, and duration and household direct medical costs and total costs (Tables 3 and 4). Both direct and total costs tend to be lower for children in the lowest wealth quintile in all 3 countries; however, with only marginal statistical significance. In Bangladesh, there is a trend toward increasing medical and total costs with wealth. In India, both types of costs take the form of an inverted U-shape, with costs increasing for the middle and fourth quintiles and then declining again for the richest. For Pakistan they are relatively consistent across wealth groups.

In Bangladesh, medical costs were higher for girls than for boys, but there are no apparent differences for total costs. All 3 countries show a trend to higher costs with higher levels of education, especially for household direct medical costs.

In India and Pakistan there is a trend toward lower direct medical costs for older children (45%–65% less), and to a lesser extent for total household costs. There is a trend for greater cost for moderate-to-severe diarrhea (45%–50% greater) in Bangladesh and India, compared to mild, but this was less marked for Pakistan. There is also a trend toward higher costs with greater duration, but the pattern is inconsistent.

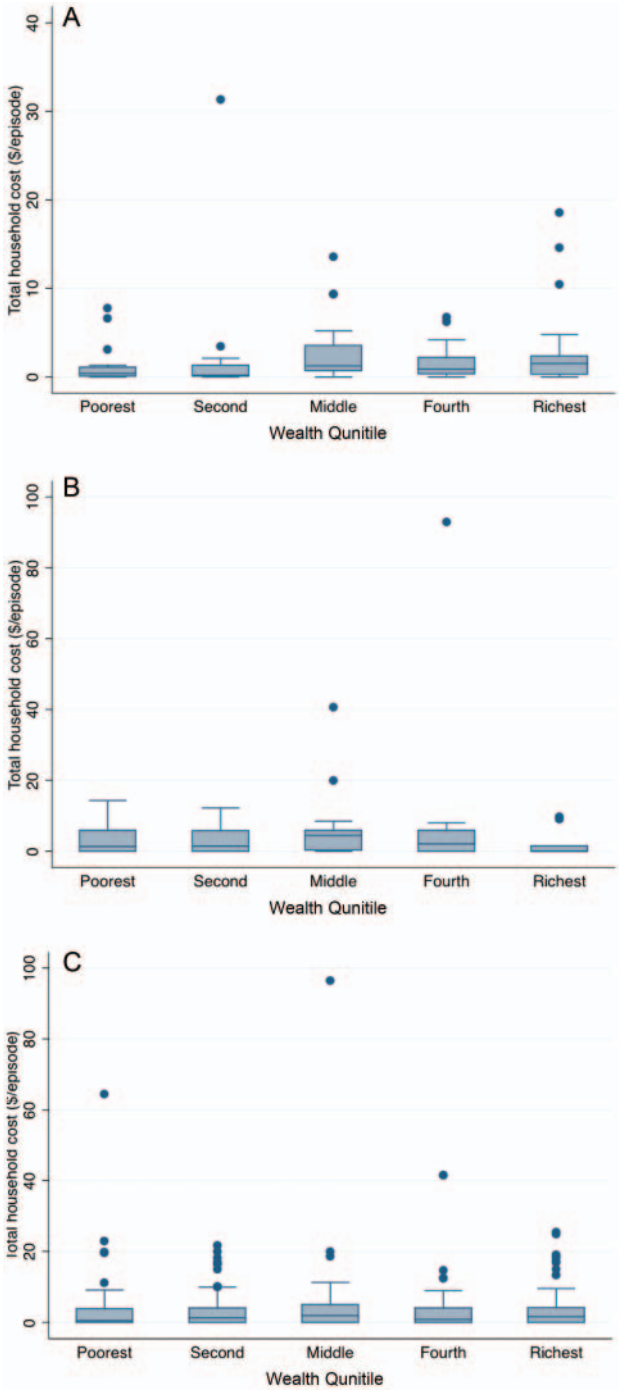

Figure 1. Distribution of total household diarrhea costs by wealth quintile ($ per episode) in (*A*) Bangladesh, (*B*) India, and (*C*) Pakistan.

Costs as a Barrier

Table 5 shows the results for questions relating to why parents did not seek care of diarrhea or why they did not seek hospital attention when it was recommended. In all 3 countries, the primary reason for not seeking care was a perception that no

Table 3. Household Direct Medical Costs for Childhood Diarrhea by Socioeconomic, Demographic, and Illness Characteristic in Bangladesh, India, and Pakistan

	Bangladesh						India						Pakistan					
	Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs	
	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA
Wealth quintile																		
Poorest	1.00		0.61		1.65		0.99		0.40		2.47		1.35		0.53		2.55	
Second	0.53	*P* = .41	0.93	.02	0.57	*P* = .41	1.92	*P* = .56	0.72	.13	2.89	*P* = .60	1.47	*P* = .08	0.68	.11	2.24	*P* = .12
Middle	1.19		0.96	.02	1.24		3.48		0.81	.02	4.85		2.81		0.65	.26	4.24	
Fourth	1.08		0.89	.06	1.21		2.64		0.59	.29	4.67		0.88		0.53	1	1.71	
Richest	1.12		0.88	.08	1.27		1.16		0.36	.84	3.26		2.95		0.57	.75	5.41	
Sex																		
Male	0.71		0.86		0.82		2.13		0.54		4.20		2.1		0.61		3.61	
Female	1.30	*P* = .13	0.85	.92	1.53	*P* = .13	2.03	*P* = .46	0.60	.67	3.54	*P* = .50	1.53	*P* = .44	0.58	.64	2.65	*P* = .36
Education																		
None to primary	0.66		0.83		0.8		1.20		0.50		2.64		1.32		0.6		2.2	
Finished primary	1.04	*P* = .20	0.84	.87	1.23	*P* = .29	3.20	*P* = .16	0.70	.17	4.74	*P* = .49	2.44	*P* = .08	0.72	.31	3.53	*P* = .01
Some secondary	1.18		0.95	.22	1.24		1.61		0.41	.63	3.92		3.71		0.62	.86	5.98	
Religious only	…		…		…		3.54		1.00		3.54		2.58		0.53	.37	5.33	
Age																		
0–11 mo	1.01		0.9		1.13		4.80		0.71		6.97		2.77		0.64		4.36	
12–23 mo	1.14	*P* = .69	0.82	.32	1.39	*P* = .56	1.51	*P* = .09	0.65	.61	2.80	*P* = .22	1.51	*P* = .10	0.56	.2	2.72	*P* = .22
24–35 mo	0.77		0.85	.6	0.9		1.40		0.50	.08	2.82		1.49		0.59	.49	2.64	
Severity																		
Mild	0.53		0.88		0.61		1.20		0.41		3.07		1.34		0.44		2.89	
Moderate/severe	1.07	*P* = .07	0.85	.68	1.27	*P* = .08	2.33	*P* = .25	0.62	.16	3.97	*P* = .50	1.89	*P* = .53	0.62	.04	3.17	*P* = .74
Duration																		
1–3 d	0.70		0.85		0.83		1.54		0.57		2.86		2.27		0.71		3.31	
4–7 d	1.13	*P* = .27	0.9	.59	1.25	*P* = .41	3.27	*P* < .001	0.86	.09	4.03	*P* < .001	1.87	*P* < .001	0.74	.71	2.66	*P* < .001
8–14 d	0.78		0.92	.56	0.85		11.64		0.26	.31	44.01		1.42		0.5	.19	2.82	
≥15 days	2.34		1		2.34		6.53		1.00		6.53		8		0.64	.67	12.58	

Table 4. Household Total Costs for Childhood Diarrhea by Socioeconomic, Demographic, and Illness Characteristic in Bangladesh, India, and Pakistan

	Bangladesh						India						Pakistan					
	Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs		Mean Cost for All Episodes		Proportion With Costs		Mean Cost for Episodes With Costs	
	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA	$	ANOVA		*P* Value	$	ANOVA
Wealth quintile																		
Poorest	1.36		0.61		2.24		2.38		0.50		4.73		4.08		0.53		7.71	
Second	1.22	*P* = .64	0.93	.02	1.31	*P* = .73	3.65	*P* = .74	0.72	.30	5.10	*P* = .83	3.40	*P* = .86	0.78	.01	4.36	*P* = .77
Middle	2.39		0.96	.02	2.49		5.07		0.81	.07	6.25		3.94		0.69	.14	5.71	
Fourth	1.89		0.89	.06	2.11		3.63		0.66	.36	5.46		4.35		0.64	.28	6.72	
Richest	2.68		0.88	.08	3.04		1.38		0.36	.52	3.87		4.19		0.56	.81	7.39	
Sex																		
Male	1.51		0.86		1.76		3.17		0.60		5.30		4.42		0.66		6.73	
Female	2.31	*P* = .87	0.85	.92	2.72	*P* = .84	3.47	*P* = .57	0.65	.67	5.32	*P* = .58	3.50	*P* = .16	0.62	.60	5.56	*P* = .22
Education																		
None to primary	1.71		0.83		2.07		2.49		0.56		4.47		3.79		0.65		5.79	
Finished primary	1.44	*P* = .17	0.84	.87	1.70	*P* = .20	4.69	*P* = .23	0.76	.15	6.18	*P* = .51	4.91	*P* = .54	0.72	.57	6.76	*P* = .48
Some secondary	3.26		0.95	.22	3.42		1.72		0.41	.42	4.17		4.81		0.62	.78	7.75	
Religious only							5.25		1.00		5.25		3.97		0.58	.36	6.73	
Age																		
0–11 mo	2.74		0.90		3.04		6.98		0.74		9.47		4.30		0.70		6.14	
12–23 mo	2.09	*P* = .37	0.82	.32	2.55	*P* = .42	2.56	*P* = .14	0.65	.47	3.92	*P* = .24	2.91	*P* = .33	0.61	.14	4.68	*P* = .43
24–35 mo	1.14		0.85	.60	1.34		2.42		0.58	.20	4.16		4.33		0.63	.31	6.90	
Severity																		
Mild	1.06		0.88		1.20		1.67		0.41		4.06		1.86		0.49		3.57	
Moderate/severe	2.08	*P* = .27	0.85	.68	2.45	*P* = .31	3.81	*P* = .20	0.69	.06	5.52	*P* = .47	4.29	*P* = .14	0.66	.05	6.46	*P* = .23
Duration																		
1–3 d	1.60		0.85		1.89		2.49		0.60		4.12		4.92		0.72		6.82	
4–7 d	2.19	*P* = .63	0.90	.59	2.43	*P* = .70	4.29	*P* < .001	0.86	.13	4.98	*P* < .001	3.44	*P* < .001	0.78	.43	4.41	*P* < .001
8–14 d	0.83		0.92	.56	0.90		19.04		1.00		19.04		3.94		0.66	.73	5.97	
≥15 days	3.06		1.00		3.06		8.48		1.00		8.48		17.87		0.78	.62	21.69	

Table 5. Reasons for Not Seeking Treatment and Sources of Household Costs for Diarrhea Episodes in Bangladesh, India, and Pakistan

Question	Pakistan (n = 68)	Bangladesh (n = 11)	India (n = 20)
Why did households not seek care for their child?			
No need for care	30.9%	72.7%	95.0%
Distance too far	4.4%	9.1%	0.0%
Lack of transportation	1.5%	0.0%	0.0%
No time off work	11.8%	18.2%	5.0%
Local situation (political)	2.9%	0.0%	0.0%
Transportation costs	5.9%	0.0%	0.0%
Treatment costs	27.9%	18.2%	5.0%
Leaving other children at home	5.9%	0.0%	5.0%
Unhappy with clinical services	11.8%	0.0%	0.0%
Prefer traditional medicine	5.8%	0.0%	0.0%
Cultural differences	0.0%	0.0%	0.0%
Other	19.1%	0.0%	0.0%
Why did the household not seek hospital care when advised?	Pakistan (n = 19)	Bangladesh (n = 14)	
Hospital too far	0.0%	14.3%	
No transportation	0.0%	0.0%	
Travel costs too high	5.3%	7.1%	
No time off work	0.0%	7.1%	
Local situation	0.0%	0.0%	
Treatment costs	47.4%	21.4%	
Needs of other children at home	21.1%	0.0%	
Child not sick enough	0.0%	42.9%	
Unhappy with clinical services	15.8%	0.0%	
Other	10.5%	7.1%	
Where did the money come from?	Pakistan (n = 281)	Bangladesh (n = 81)	India (n = 62)
Fewer meals	12.5%	1.2%	19.4%
Cutting other expenses	13.9%	1.2%	21.0%
Savings	47.7%	80.3%	48.4%
Borrowing	29.2%	9.9%	21.0%
Selling assets	3.2%	4.9%	3.2%
Donations	1.4%	0.0%	0.0%
Relative or friend	3.9%	2.5%	1.6%
Other	7.8%	4.9%	1.6%

care was needed, with 31%, 73%, and 95% of parents reporting this in Pakistan, Bangladesh, and India respectively. Among the remainder who thought that care was needed, the main causes related directly (eg, transportation, treatment) or indirectly (eg, no time off work, lack of transportation, and distance) to costs. Among those who were recommended to take their child to the hospital and did not, no perceived need was the primary reason in Bangladesh (43%) and costs were the primary reason in Pakistan (53%).

The main source of funding for household costs for diarrhea episodes was savings in all 3 countries, with 48% in Pakistan, 48% in India, 80% in Bangladesh reporting this. The other common sources were borrowing and cutting expenses.

DISCUSSION

What Are the Costs of Diarrhea to Households?

Few studies measure the financial cost of a diarrheal illness borne by the family and the healthcare system, yet this information becomes critical when decision makers must set priorities for designing and implementing public health interventions to prevent and treat these illnesses. The HUAS provided an opportunity to assess the direct costs and productivity losses associated with a diarrheal illness during the first 5 years of life in a random sample of children living in resource-poor settings in South Asia. These analyses, undertaken in preparation for the GEMS case/control study that will provide prospective, pathogen-specific data on the costs of moderate-to-severe medically

attended diarrheal illnesses in the same population, demonstrate in 3 Asian sites that financial costs represent an important component of the diarrheal disease burden.

The 3 countries provide different patterns of household costs from diarrhea. Both direct medical and total costs are lower in Bangladesh than in India and Pakistan (approximately 50%). Interestingly, the proportion of episodes that incur some costs is higher in Bangladesh, suggesting that households are more likely to seek care, and when they do the amount they spend is less. Among episodes that incur some costs, the total cost is less in Bangladesh compared with India or Pakistan.

This difference is almost completely accounted for by the difference in direct medical costs, with little difference in other direct costs or indirect costs. This is also reflected in where care is sought. In Bangladesh approximately half of the household direct medical cost is for care in informal settings, while in India and Pakistan 87% and 75% of costs are in the formal sector. This may reflect higher utilization of low-cost oral rehydration from pharmacies and other outlets. While the lower level of costs in Bangladesh may reflect a lower income level for the setting, it is also likely to be influenced by the healthcare system and high awareness of early treatment of diarrhea with oral rehydration therapy.

Average household costs provide only one aspect of the economic burden. Given the high frequency of diarrhea among young children in low-income settings, there is a possibility that a low-probability but high-cost episode might occur, creating a financial strain for the household. Based on the results presented here, for each diarrhea episode a household faces, there is a 1 in 10 chance of a total cost greater than $6.61 in Bangladesh, $8.07 in India, and $10.11 in Pakistan. In countries where many households live on less than $1 per day, this represents a substantial risk.

In all 3 countries there is some evidence of lower costs for children in households with lower economic or educational levels. This is consistent with poorer households being more likely to ration or delay care due to high costs. This brings with it the risk that delayed treatment will result in more severe outcomes. Although we are not able to address this directly in this study, it should be addressed empirically in subsequent analyses.

Costs by sex differed between countries, with high costs for girls in Bangladesh and higher costs for boys in Pakistan, and little difference in India. This was true for direct medical and total costs. This deserves additional exploration given the evidence in the literature that household expenditures for healthcare, food, and education often favor boys over girls. [22–25]. In making decisions about whether and when to invest in treatment, households may be willing to take greater risks with girls. It is important to note that household costs for girls are not lower in Bangladesh, the where overall household economic burden is less and there may be less need to ration care.

Cross-Country Comparisons

The present study in combination with our related study on household costs for diarrhea treatment in African settings provides helps identify similarities and differences among countries and regions [26]. Across the 6 countries, mean total household costs fell within a fairly wide range, from $1.82 in Bangladesh to $6.47 in Pakistan. In general, costs were higher in higher income settings, reflecting available household resources and the cost of services. However, mean out-of-pocket costs also appear to reflect health system characteristics, in particular the level of subsidy for direct medical costs. In most countries, costs tended to be lower among low-income households, potentially reflecting rationing of care; however, others exhibited an inverted U-shape with the highest costs in the middle-income levels. As described above, 3 of the 6 countries demonstrated lower costs for girls; however, this was not present in others. A number of factors may affect whether there are observed differences between boys and girls. Household income and the absolute level of costs borne by households may affect the need to ration care.

Limitations

The current work suffers from several important limitations. First, the study sample size was designed to provide estimates of overall costs within a margin of error. This is particularly true for India and Bangladesh. It was not powered to examine the determinants of costs and as a result some differences in means among subgroups are often not statistically significant. Additional work with larger samples would help address this. Second, with one-time cross-sectional data we are not able to directly examine the long-term health or economic consequences of the costs incurred by household for individual events or repeated episodes. Lastly, the cross-sectional nature of the study makes it difficult to assess whether low costs for specific subgroups are the result of reduced severity, cheaper services, or rationing of care. Additional work must also be conducted to better understand how the complex interaction between direct medical, direct nonmedical, and indirect costs impact households' demand for and decisions to seek informal or formal care.

CONCLUSIONS

While the absolute value of household economic costs are relatively low for each childhood diarrhea episode, their cumulative impact is likely to be great. Given the frequency of diarrheal episodes among children in low-income settings, these small amounts per case would translate into billions of dollars globally, borne by the families themselves. The small average costs also hide the repeated possibility that an episode will require more extensive and expensive care, resulting in

indebtedness or selling of productive assets. Our results also suggest that the household economic burden may result in some households choosing to reduce or delay direct medical expense, especially poor households or for girls. This suggests that these costs and other barriers to care may create or accentuate disparities in adverse outcomes including mortality.

Notes

Financial support. This work was supported by the Bill & Melinda Gates Foundation (grant number 38874).

Supplement sponsorship. This article was published as part of the supplement entitled "The Global Enteric Multicenter Study (GEMS)," sponsored by the Bill & Melinda Gates Foundation.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet **2010**; 375:1969–87.
2. Atherly DE, Lewis KD, Tate J, Parashar UD, Rheingans RD. Projected health and economic impact of rotavirus vaccination in GAVI—eligible countries: 2011–2030. Vaccine **2012**; 30(suppl 1):A7–14.
3. Chai PF, Lee WS. Out-of-pocket costs associated with rotavirus gastroenteritis requiring hospitalization in Malaysia. Vaccine **2009**; 27(suppl 5):F112–5.
4. Fischer TK, Anh DD, Antil L, et al. Health care costs of diarrheal disease and estimates of the cost-effectiveness of rotavirus vaccination in Vietnam. J Infect Dis **2005**; 192:1720–6.
5. Forsberg BC, Sullesta E, Pieche S, Lambo N. Costs of diarrhoeal diseases and the savings from a control programme in Cebu, Philippines. Bull World Health Organ **1993**; 71:579–86.
6. Jeuland M, Whittington D. Cost–benefit comparisons of investments in improved water supply and cholera vaccination programs. Vaccine **2009**; 27:3109–20.
7. Podewils LJ, Antil L, Hummelman E, Bresee J, Parashar UD, Rheingans R. Projected cost-effectiveness of rotavirus vaccination for children in Asia. J Infect Dis **2005**; 192(suppl 1):S133–45.
8. Rheingans R, Atherly D, Anderson J. Distributional impact of rotavirus vaccination in 25 GAVI countries: estimating disparities in benefits and cost-effectiveness. Vaccine **2012**; 30(suppl 1):A15–23.
9. Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski GE, ed. Changing the US health care systems: key issues in health services policy and management. San Francisco: Jossey-Bass, **2001**:3–30.
10. Chola L, Robberstad B. Estimating average inpatient and outpatient costs and childhood pneumonia and diarrhoea treatment costs in an urban health centre in Zambia. Cost effectiveness and resource allocation: C/E **2009**; 7:16.
11. Gelberg L, Andersen RM, Leake BD. The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. Health Serv Res **2000**; 34:1273–302.
12. Banerjee A, Deaton A, Duflo E. Health, health care, and economic development: wealth, health, and health services in rural Rajasthan. Am Econ Rev **2004**; 94:236–30.
13. Stanton B, Clemens J. User fees for health care in developing countries: a case study of Bangladesh. Soc Sci Med **1989**; 29:1199–205.
14. Robberstad B, Strand T, Black R, Sommerfelt H. Cost-effectiveness of zinc as adjunct therapy for acute childhood diarrhoea in developing countries. Bull World Health Organ **2004**; 82:523–31.
15. World Health Organization. Guidelines for estimating the economic burden of diarrheal disease with focus on assessing the costs of rotavirus diarrhea. Geneva, Switzerland: WHO, **2005**.
16. Mwabu G, Mwanzia J, Liambila W. User charges in government health facilities in Kenya: effect on attendance and revenue. Health Policy Plan **1995**; 10:164–70.
17. Nyman JA. Is 'moral hazard' inefficient? The polity implications of a new theory. Health Aff **2004**; 23:194–9.
18. Patil AV, Somasundaram KV, Goyal RC. Current health scenario in rural India. Aust J Rural Health **2002**; 10:129–35.
19. Rao NS. Delivery of health services in the public sector. Financing and delivery of health care services in India. In: National Commission on Macroeconomics and Health Background Papers. New Delhi: Ministry of Health and Family Welfare, **2005**:43–59.
20. Levine MM. Etiology of diarrhoeal and enteric diseases in developing countries: the Global Enteric Multicenter Study (GEMS) project. Global Vaccine Research Forum. Bamako, Mali: WHO, **2009**.
21. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. Measurement of living standards. Analyzing health equity using household survey data: a guide to techniques and their implementation. Washington, DC: World Bank, **2008**:69–82.
22. Nguyen H, Knowles J. Demand for voluntary health insurance in developing countries: the case of Vietnam's school-age children and adolescent student health insurance program. Soc Sci Med **2010**; 71:2074–82.
23. Nuruddin R, Hadden WC, Petersen MR, Lim MK. Does child gender determine household decision for health care in rural Thatta, Pakistan? J Public Health (Oxf) **2009**; 31:389–97.
24. Pandey A, Sengupta PG, Mondal SK, et al. Gender differences in healthcare-seeking during common illnesses in a rural community of West Bengal, India. J Health Popul Nutr **2002**; 20:306–11.
25. Pokhrel S, Snow R, Dong H, Hidayat B, Flessa S, Sauerborn R. Gender role and child health care utilization in Nepal. Health Policy **2005**; 74:100–9.
26. Rheingans R, Kukla M, Adegbola R, et al. Exploring household economic impacts of childhood diarrheal illnesses in three African settings. Clin Infect Dis **2012**; 55(S4):S317–26.

Clinical Infectious Diseases

http://cid.oxfordjournals.org/

An Official Publication of the Infectious Diseases Society of America and the HIV Medicine Association