The Epidemiology of Rotavirus Diarrhea in the United States: Surveillance and Estimates of Disease Burden

Roger I. Glass, Paul E. Kilgore, Robert C. Holman, Shaoxiong Jin, Jean C. Smith, Patricia A. Woods, Matthew J. Clarke, Mei Shang Ho, and Jon R. Gentsch

The decision to develop rotavirus vaccines was predicated on the extensive burden of rotavirus disease among children worldwide. US reports on nationwide hospitalizations (1979–1992) and deaths (1968–1991) due to diarrhea and weekly reports of rotavirus infection by 74 laboratories were reviewed to estimate the burden of rotavirus disease, identify epidemiologic trends, and consider methods for evaluating an immunization program when a vaccine becomes available. From 1968 to 1985, diarrhea-related deaths among US children <5 years old declined from 1100 to 300/year. This decline was associated with the disappearance of winter peaks for diarrhea-related deaths previously associated with rotavirus infection among children 4–23 months old. From 1979 to 1992, however, hospitalizations for diarrhea averaged 186,000/year and retained their winter peaks, which have been linked to rotavirus infections. Each year an estimated 54,000–55,000 US children are hospitalized for diarrhea, but <40 die with rotavirus. A rotavirus vaccine program will require improved surveillance, including the timely collection of data from sentinel hospitals, in which a diagnosis of rotavirus can be established or ruled out for all children hospitalized for diarrhea.

The need for a rotavirus vaccine for children in developing countries has long been evidenced by the tremendous burden of rotavirus diarrhea [1, 2]. Many epidemiologic studies have documented the high prevalence of severe rotavirus diarrhea among children in developing countries, accounting for, on average, one-third of all hospitalizations for diarrhea worldwide [3] and an estimated 873,000 deaths/year, which is equivalent to 25% of all diarrheal deaths and 6% of all deaths in children <5 years old. In 1985, similar information about diarrheal disease was not available for developed countries, so the Institute of Medicine, in reviewing new vaccine priorities for developed countries, concluded that a rotavirus vaccine was not a priority for the United States [4]. The decision was based on the results of a prospective study of the incidence of hospitalizations for rotavirus diarrhea in a birth cohort of 126 children followed for 2 years [5] and not on hospital-based studies that had already documented that one-third of children admitted for diarrhea have rotavirus infection [6]. While most rotavirus diarrhea in American children is mild, nearly every child is infected in the first 2–3 years of life, and a percentage of these episodes are severe enough to warrant a visit to a doctor or hospitalization. By 1985, death, the most severe outcome of rotavirus infection, had not been documented in a child in the United States; better data were needed to quantify the more severe and costly outcomes of rotavirus.

In 1987 in response to this challenge, we began a systematic study of the epidemiology of rotavirus diarrhea in the United States. The study was based largely upon mortality and hospitalization data that were freely available and in the public domain [7, 8]. Data also included some from the Rotavirus Surveillance Network, an initial group of 88 North American laboratories that were recruited to voluntarily report rotavirus detections on a weekly basis to the Centers for Disease Control and Prevention (CDC) [9]. Our original goals were to understand the epidemiology of rotavirus in the prevaccine era, to examine the disease burden of rotavirus diarrhea, and to identify an indicator(s) of the disease in the epidemiologic data that were available. For hospitalizations and deaths, the task was a challenge, because no code from the ninth revision of the International Classification of Diseases (ICD-9) existed for rotavirus; we had to work with a collection of nonspecific codes for diarrhea (e.g., diarrhea, presumed noninfectious) and identify epidemiologic patterns that would specifically define rotavirus disease. Besides providing retrospective information on diarrheal hospitalizations and deaths in the United States, we felt that this survey would help define methods to assess the disease burden of rotavirus in other countries and to gain experience with the feasibility of different surveillance methods to monitor the impact of a rotavirus vaccine program when such a program was implemented.

This paper reviews studies done primarily by the CDC that laid the groundwork for our appreciation that rotavirus diarrhea is an important cause of disease leading to hospitalization of children in the United States. As a result of these studies and others, we now have national measures of both morbidity and mortality to estimate the disease burden of rotavirus diarrhea. These data have been used to calculate the cost-effectiveness...
of a rotavirus vaccine program, identify unusual epidemiologic features of rotavirus, such as the west-to-east dissemination pattern of the disease, and document that the peak of winter deaths in the United States attributed to rotavirus has declined in recent years. From these activities, we have further identified areas where improved surveillance could provide timely data to assess the impact of an intervention program with vaccines.

Methods and Findings

Background. To estimate the disease burden of rotavirus diarrhea, we quantified all clinically severe diarrheal events (i.e., deaths, hospitalizations, and doctor visits) and tried to identify, on the basis of our understanding of the epidemiology of the disease, those events that might be due to rotavirus. We turned to two national sources of data available from the National Center for Health Statistics (NCHS, CDC), the multiple cause-of-death data [10] and the National Hospital Discharge Survey (NHDS) data [11], and to data on the laboratory surveillance of rotavirus that is linked with the National Enteric and Respiratory Surveillance System [12]. Since no codes in the ICD-9 or ICD-9-Clinical Modification (ICD-9-CM) were available for rotavirus diarrhea before 1993, we had to use a collection of codes for diarrhea, even though we knew that >90% of diarrheal events, even those leading to hospitalization or death, are coded without an etiologic diagnosis [13, 14].

For analysis of diarrhea mortality, we included reports of deaths that had diarrhea coded anywhere on the NCHS death record. For analysis of hospitalizations, we included records in which codes for diarrhea were listed in the first three positions on the discharge record. Previous experience had shown us that diarrhea would be listed as the first code on the record for 73% of the hospitalizations and as the third code for 95% [15]. Use of records with diarrhea recorded only as the first cause of hospitalization seemed restrictive, since dehydration, a related diagnosis, was often reported as the first cause of hospitalization, and limiting diarrheal diagnosis to the top three positions would exclude only 5% of key events that might be secondary to the true cause of hospitalization.

NHDS data. Information on hospital discharges has proven to be the most robust, sensitive, and important data available to monitor rotavirus activity in the United States. Events reported in these discharge reports are costly (>1000/day) and numerous, and our experience indicates that they not only reflect the level of rotavirus activity but reveal epidemiologic trends that were previously unrecognized [7]. Earlier etiologic studies of children hospitalized for diarrhea in the United States suggested that hospital discharge would be a strong predictor of rotavirus activity: rotavirus-related diarrhea was the most common cause of hospitalization for diarrhea among children <5 years (~30%), rotavirus activity was highly seasonal (winter months), and rotavirus diarrhea was limited to children between 3 and 35 months of age [6, 16–18].

The first rotavirus studies by Ho et al. [7, 8] used NHDS data from 1979-1984 with a variety of ICD-9—CM codes for diarrhea, most of which were nonspecific, to look for national trends in diarrheal hospitalizations by age, time, and geographic region that might reflect rotavirus activity. The NHDS data represent a systematic sample of patient discharges obtained from hospitalized short-stay non-federal general and specialty hospitals in the United States [18a]. To look more carefully at age, geographic, and seasonal trends that required a larger data set, Ho and colleagues [8] examined hospital discharge data from the McDonnell-Douglas Automated Data Information Systems, which represented a 16% nonsystematic sample of all hospitalizations in the United States [7]. Results from the two surveys were comparable, and the larger data set only helped to decrease the confidence limits of the smaller data set: ~210,000 children 1–59 months of age were hospitalized each year for diarrhea, representing ~9.5% of all hospitalizations among children of this age group. These hospitalizations peaked from November through April, and these peaks of activity were primarily among children 3–35 months of age. Moreover, the peak of this activity began in the Southwest in November and moved across the United States, ending in the Northeast in March and April. A 5-year national survey of the detection of rotavirus by 88 US laboratories was done by LeBaron et al. [9]. This study confirmed that the peak of rotavirus detections followed the same geographic and temporal patterns as hospitalizations and deaths for diarrhea and suggested that the winter seasonal peaks of hospitalizations for diarrhea were due to rotavirus.

Since the original rotavirus study (Ho et al. [7]) ended with data from 1984, Jin et al. [19] repeated and updated the study with data from 1979 to 1992, the last year for which hospital discharge data were available, to see if hospitalization for diarrhea, particularly in the winter, remains a problem (figure 1A). After reviewing information on 16,049 hospitalizations, representing a weighted total of 2.6 million discharges (i.e., annual average of 186,000 hospitalizations for diarrhea/year), their findings were comparable to those of Ho et al. [7] and others. While the annual number of diarrheal hospitalizations decreased by 13% over time, the winter seasonal peaks remained constant, and the geographic and temporal appearance of these peaks—first in the Southwest and later in the Northeast—was unchanged. Jin et al. [19] found that the winter peak for diarrheal diseases coded as “viral” (19% of the total and many of which are presumed to be due to rotavirus) was mimicked by the larger number of diarrheal diseases that were presumed to be noninfections (71% of the total) and by the total group of diarrheal episodes, suggesting that these groups shared a common etiology. Only the small number (5%) of diarrheal episodes caused by bacteria had a different seasonal pattern.

In the absence of hard data on laboratory-confirmed cases of rotavirus diarrhea, we developed two distinct and independent methods to estimate the proportion of hospitalizations that might be attributable to rotavirus [19, 21]. The first method,
Figure 1. Hospitalizations (A) and deaths (B) due to diarrhea by month among US children <5 years old, 1979-1992. Reprinted with permission from Jin et al. [19] and Kilgore et al. [20]. Mo = months; Yr(s) = year(s).
the residual estimate, defined rotavirus-related hospitalizations as the excess number of hospitalizations in winter versus summer season. Summer was defined as the 6-month period without rotavirus activity (May–October); winter included the months November–April. Since the summer baseline would include bacterial diarrheas not present in the winter, we felt this estimate would be conservative, possibly underestimating the percentage of the excess peak due to rotavirus. The excess was greatest during the month of January and was least in November and April, representing the geographic hospitalization peaks in the Southwest and the Northeast, respectively.

The second method, rate standardization estimate, was derived by multiplying the month-specific rate of rotavirus detection among children hospitalized for diarrhea over an 8-year period in Washington, DC, by the monthly rates of diarrheal hospitalization in the United States over the 14-year period taken from the discharge survey [6, 19]. While other hospital-based surveillance studies have been done, this study is the largest, and the data are reported by month, allowing proper standardization of rates. Furthermore, these rates do not differ greatly from those of other smaller hospital-based studies. When these two estimates were compared, they yielded time curves that were synchronous and numeric predictions of the same magnitude—about 55,000 hospitalizations/year.

National mortality data. Since no child in the United States had ever been reported to have died with a diagnosis of rotavirus diarrhea, many pediatricians have felt that it is never severe or fatal. A review of national mortality data, initially for 1972–1983 and more recently for 1968–1991, has provided cogent but indirect evidence that rotavirus has been a cause of death in American children (figure 1B) [8, 20]. The NCHS’s multiple cause-of-death data were examined to extract diarrheal deaths with a variety of ICD codes from the adapted ICD-8 (1968–1978) and ICD-9 (1979–1991) [13, 22]. Most of the deaths (>90%) had no etiologic diagnosis, and since no ICD code was available for rotavirus during this study period, none of the deaths could be attributed directly to rotavirus.

Nonetheless, the analysis of data for 14,137 deaths due to diarrhea that occurred from 1968 through 1991 provided some insight into the role that rotavirus diarrhea may play as a cause of death in American children [20]. In the United States, the number of deaths due to diarrhea has declined continuously from ~1100 deaths/year in 1968 to ~300/year for each year from 1985 to 1991. These deaths had a distinct winter seasonal peak that was most marked among children 3–24 months old, epidemiologic characteristics similar to those observed for children hospitalized for diarrhea. An unexpected finding of this study was that this peak showed the same geographic and temporal trend as hospitalizations and laboratory detections of rotavirus beginning in the Southwest in November and ending in the Northeast in March–April. Most (78%) of these deaths were in children <1 year old. Diarrhea appeared to be the likely cause of death, since other diagnoses were most often problems of electrolyte imbalance, dehydration, and other complications of diarrhea.

The decline in deaths from 1968 to 1985 was associated with a marked dampening of the winter peaks for diarrheal disease. Since 1985, an excess of winter deaths associated with diarrhea is still evident in children 3–23 months old, but it is much less prominent than before. Ho et al. [8] originally attributed the excess of winter versus summer deaths due to diarrhea to rotavirus and determined they represented 75–125 deaths/year [8]. According to data for the most recent years for which data on deaths due to diarrhea are available, 1985–1991, these deaths now number about 20/year [20].

We hypothesized that this decline in winter deaths attributable to rotavirus could be due either to a decrease in the overall incidence of severe diarrhea in winter or to improvements in the case management of diarrhea or access to care. The unchanged number of winter hospitalizations indicates that the decline in deaths must be due to other factors, such as improved case management or access to care.

National laboratory surveillance. All data described so far have been independent of a specific diagnosis of rotavirus. The observation that the peak of winter hospitalizations and deaths moved from the Southwest in November toward the Northeast in March–April led us to establish the Rotavirus Surveillance Activity to monitor this pattern prospectively for 1985–1989 and prospectively to the present [9, 12, 23]. The 5-year retrospective survey confirmed for the first time the unusual geographic and temporal trends of rotavirus detections, which represent unique epidemiologic indicators for the disease. The fact that this pattern occurs in a repeated fashion undermines simple explanations that temperature or humidity caused the seasonal peaks of rotavirus. The alternative hypothesis, that individual strains were migrating, was also dismissed when Gouvea et al. [24] found multiple strains emerging in different parts of the United States in the same season and when strains from the same city in different seasons were often similar, suggesting a common local reservoir. Similar observations were made by Matson et al. [25].

National reporting of rotavirus detections from 74 laboratories in the United States has been ongoing since the original survey by LeBaron et al. [9]. The national surveillance has identified summer peaks of rotavirus activity that, when investigated, were found to be due to poor laboratory practices (i.e., >6 different technicians doing the test) and to the use of a commercial test that was sold without proper controls [23]. While reporting is weekly and on a voluntary basis, much information that would be important for evaluating an intervention program is not now collected. For instance, the laboratories do not record the patient’s age or underlying illness or whether patients were hospitalized or seen as outpatients or whether the infection was acquired in the community or in the hospital. In addition, specimens are not submitted, so there can be no assessment of rotavirus strains currently in circulation.
Estimates of disease burden. A major goal of these epidemiologic studies has been to quantify the role played by rotavirus as a cause of diarrhea among children in the United States. Early longitudinal studies and recent placebo-controlled field trials have indicated that if surveillance is conducted actively with two or more home visits per week, the incidence of rotavirus diarrhea may reach 0.30 episodes/child/year during the first 2 years of life, with a cumulative incidence approaching 0.80 episodes in a child reaching age 5 [5, 26-28]. Doctor visits are less frequent (~1/8 children) by age 5, but the national database year from which this estimate is derived is the least robust and has the greatest margin of error [29]. On the basis of NCHS hospitalization data for all diagnoses of diarrhea, we estimate that 1 in 25 children will be hospitalized for diarrhea during the first 5 years of life (as determined by the estimate of 186,000 hospitalizations for diarrhea in children 1-59 months old in a birth cohort of 4 million children followed for 5 years). About 1 in 73 infants will be hospitalized for rotavirus diarrhea, as determined using our estimate of 55,000 hospitalizations. The 20 deaths per year (1 per 200,000 live births) probably represent a subgroup of children with underlying illnesses or conditions, such as prematurity [20].

These data on the national disease burden of rotavirus diarrhea must be placed in context of the other estimates that have been made (table 1). First, the Institute of Medicine in 1985 estimated the number of hospitalizations for rotavirus at 23,000 per year by extrapolating data from a longitudinal study of 126 children to the total population of children in the United States [4]. Second, Ho et al. [7] estimated 65,000-70,000 hospitalizations/year by using the approach described in this paper with hospital discharge data from 1979-1985. Third, Matson et al. [17] used data for rotavirus hospitalizations at Texas Children’s Hospital to prepare broader estimates of 110,000 hospitalizations/year, one-third of which were associated with nosocomial infections. Last, Smith et al. [30] estimated 104,000 rotavirus-related hospitalizations/year by updating the figures of Ho et al. with newer information on the growth of the birth cohort over the intervening decade. The current figure is most similar to that of Ho et al., and the decrease in the total number of hospitalizations reflects a 13% decline in the incidence of all hospitalizations for diarrhea over the past decade.

Surveillance of laboratory strains. A number of surveys have been done to examine rotavirus strains collected in the United States [24, 25, 31, 32]. Rotavirus strains are characterized by the outer capsid proteins that are important for virus neutralization: glycoprotein (G protein) encoded by gene segment 7, 8, or 9 and the protease-cleaved protein (P protein) encoded by gene segment 4. The surveys indicate that only 4 strains are commonly circulating—P[8], G1, P[4], G2, P[8], G3, and P[8], G4—although some unusual strains have been documented on rare occasions and confirmed when both G and P typing have been done [31, 33, 34]. Strains appear to have a local reservoir, since the same strains are found from year to year in the same location. Rotavirus infects older children and adults routinely, and one might expect periodic outbreaks with illness in older children or adults if strains with major antigenic differences were to emerge. Clearly, surveillance of laboratory strains will have to continue, with specific emphasis on the identification of new circulating strains and examination of their prevalence in the population.

Discussion

In the prevaccine era, the purpose of surveillance has been limited to describing trends in rotavirus activity and obtaining better estimates of the burden of rotavirus disease. Using national data for hospitalizations and deaths and additional information from laboratory surveillance, we have been able to infer that the winter peaks of hospitalizations for diarrhea and deaths in children 3-23 months of age that appear first in the Southwest in November and move to the Northeast by March-April are attributable to rotavirus. This finding has permitted us to estimate the magnitude of the disease burden of rotavirus hospitalizations (55,000/year) and deaths (20-40/year) in the United States and assess the annual cost of this burden (>1 billion dollars) while identifying unusual seasonal and geographic patterns that represent a unique epidemiologic indicator of this disease. When we first examined diarrhea mortality in Ameri-
can children through the year 1983, we were surprised to find hundreds of deaths were occurring each year and that they had a seasonal and geographic pattern consistent with rotavirus disease. Since 1985, the total number of deaths due to diarrhea has leveled off at ~300/year, and the annual winter peak of excess deaths in children 3–23 months of age attributable to rotavirus has nearly disappeared. The observation that these winter deaths have decreased while hospitalizations have continued (i.e., an improved death-to-hospitalization ratio) is consistent with the hypothesis that these improvements are due either to better access to care or improved treatment of children with severe rotavirus diarrhea. Of note, this suspected improvement in rotavirus mortality does not markedly alter the cost-effectiveness results since most of the medical costs in the analysis are due to hospitalizations [30]. The number of hospitalizations has declined only slightly over the past 14 years, and the winter seasonal peak in children 3–35 months of age remains a distinguishing feature.

A key epidemiologic feature of rotavirus disease, which was identified in our initial analyses of deaths and hospitalizations and later confirmed in the surveillance of laboratory detection of rotavirus, is the movement of the rotavirus season in time and place, starting in the Southwest in November and moving to the Northeast in March–April. This movement indicates that neither temperature nor humidity alone can explain the trend. Laboratory studies suggest that each community may have its own local reservoir of strains, since there is not a single epidemic strain, and we have no credible hypothesis to explain this unique trend.

Surveillance of laboratory detections of rotavirus infections has provided information on rotavirus activity in the United States that is inexpensive (voluntary reporting) and timely (weekly reporting); however, it has not provided additional information that could be of considerable epidemiologic usefulness (e.g., a breakdown of infections by patient age or status [inpatient vs. outpatient] or whether the infection was community- or nosocomially acquired). As we move closer to considering implementation of a rotavirus vaccine program, sentinel surveillance will have to be established using indicator hospitals, laboratory-confirmed illness, and patients who are specifically hospitalized for diarrhea, preferably at centers with a defined population base. While laboratories are willing to voluntarily report detections of rotavirus infections, the large-scale collection and handling of rotavirus specimens for study of strain variation or the emergence of unusual serotypes will require a larger infusion of resources. In the interim, smaller surveys have confirmed that the 4 main serotypes found in repeated studies globally are also predominant in the United States.

The determination of disease burden is fraught with potential bias due to assumptions in estimating prevalence rates for rotavirus disease. We are confident that our estimates of the number of hospitalizations due to rotavirus infections are reasonable: They were determined by two completely different and independent approaches, yet the estimates are nearly identical. They are midway between the initial low estimates of the Institute of Medicine [4] and the highest estimates of Matson et al. [17] and within 15% of similar estimates we derived 8 years ago [21], and they are based on true hospitalization figures rather than estimates using hospitalization rates adjusted for the increasing population of children <5 years old [30]. They represent our best estimate made with an absence of data on hospitalizations for laboratory-confirmed rotavirus diarrhea rather than a precise determination. Better data from longitudinal studies would require that large cohorts of children (>10,000) be followed because the frequency of hospitalizations is relatively low (estimated at 1/73 children). Since a new ICD code for rotavirus diarrhea was designated in 1993 and will be included in the 10th revision of the ICD, it is likely that in the near future we may begin having direct reports of both hospitalizations and deaths due to rotavirus diarrhea.

The approach to surveillance that uses national or sentinel data on hospital discharges and deaths has identified seasonal and age trends for diarrheal illnesses, provided estimates of rotavirus prevalence, and proved to be a simple and economic method to address the epidemiology and assess the disease burden of rotavirus diarrhea in the prevaccine era. The method should be applicable in other countries that might consider the use of a rotavirus vaccine and could provide information critical to making national decisions and establishing national priorities [35]. At the same time, NHDSs provide data with a 3-year delay and without information on laboratory-confirmed cases. In view of the impending licensure of vaccines, a different approach to surveillance will be required that will permit evaluation of the effectiveness of an immunization program. This will require more intense surveillance in which sentinel hospitals provide in-depth information on patients hospitalized for rotavirus diarrhea confirmed by laboratory detection. Such sentinel hospital surveillance in larger-scale catchment trials and on a regional or national basis will be critical to monitor a decline in winter hospitalizations for diarrhea, which would be anticipated after the introduction of a rotavirus vaccine program. At the same time, increased surveillance of strains will be essential to determine if some new serotypes escape vaccination and become predominant.

Acknowledgments

We thank the former EIS Officers and students who contributed to this ongoing effort, including Charles LeBaron, Judy Lew, Caryn Bern, Helen Ciceroello, Ray Gangarosa, Donna Ing, Suzanne Cook, David Wald, Ivan Cohen, and Tom Torok; we also thank John O’Connor and Ann Risinger for contributing to manuscript preparation and editing.
References


