Using Text Messages for Critical Real-time Data Capture in the ANISA Study

Mohammad Shahidul Islam, MSc,* Qazi Sadeq-ur Rahman, MSc,† Tanvir Hossain, MSc,† Nicholas E. Connor, MSc,† Belal Hossain, MSc,* Md. Mahmudur Rahman, MSc,* Ranjan Neogi, MCom,‡ Samir K. Saha, PhD,* and Shams El Arifeen, DrPH†

Background: The Aetiology of Neonatal Infection in South Asia (ANISA) study takes advantage of text messaging technology to record information required for randomizing the study population into a control subcohort. The text message system is also used for monitoring various study activities.

Methods: When a child-health worker registers a newborn in the study, she sends a text message to a database server containing the study identification number and newborn's age at the time of registration. For each possible serious bacterial infection case, a study physician also sends a text message to the same server with the age of the young infant at the time of illness assessment. Using this information, a computer-based algorithm randomizes the newborn into a control subcohort. Text messages are also sent to alert the study physicians and study supervisors of a possible serious bacterial infection case being referred to health-care facilities. Phlebotomists working at remote specimen collection sites send text messages to the site laboratory personnel before sending the specimens through porters.

Discussion: Real-time data entry and monitoring are challenging for any population-based study conducted in remote areas. Our text messaging system provides an opportunity to overcome this barrier where availability of data entry facilities is limited.

Key Words: text message, healthy control, ANISA, seasonality, age-matched

(*Pediatr Infect Dis J* 2016;35:S35–S38)

The Aetiology of Neonatal Infection in South Asia (ANISA) study collects blood and nasopharyngeal–oropharyngeal (NP–OP) swab specimens from young infants with community-acquired possible serious bacterial infection (pSBI). These young infants are identified through active surveillance by community-health workers (CHWs) and through referral to physicians at 5 study sites in Bangladesh, India and Pakistan. The ANISA study team has developed procedures for evaluating the strength of the epidemiological association between pathogen presence in the clinical specimens obtained from ill-appearing young infants and the risk of their illness in that community. A critical challenge is the expectation that many of the same pathogens found in sick young infants may also be detected in specimens from healthy ones although possibly at a lower rate.[1] Many of the target viruses and bacteria are also part of the normal flora in the naso-oropharynx of this age group.[2–5] Hence, in this study, we collect blood and NP–OP swab specimens from a control subcohort of enrolled healthy young infants. For selecting this subcohort, we use mobile phone text message (often referred to as "SMS") technology. Although this system was initially introduced in ANISA for selecting and enrolling the control subcohort, we have extended its use to real-time critical data capture. In this article, we describe our experience with using text messages in ANISA.

ANISA STUDY METHODS

A CHW visits every household in her study area once every 2 months to identify and register pregnant women. She follows every newborn on days 0, 2, 6, 13, 20, 27, 34, 42, 48 and 59 after birth, making a total of 10 postnatal home visits. If a CHW determines any young infant to be sick based on predefined clinical criteria, she refers the child to a specific facility for reassessment by a study physician. If the physician confirms the young infant (either referred by a CHW or self-referred) as a case of pSBI, blood and NP–OP specimens are collected after written consent from the newborn's legal guardian. The detailed methodology of the pregnancy and newborn surveillance activities for this study has been described elsewhere in this supplement.[6]

CONTROL SELECTION CRITERIA

In ANISA, control specimens are collected matching 3 features of the cases. First, we considered that ANISA field sites are geographically distinct and present different pathogen sets. Thus, each study field site was considered its own stratum, and we planned to collect equal numbers of controls from each study site. Young infant infection and mortality data from South Asia, as well as historical data from industrialized countries, show a disproportionately high incidence of early-onset versus late-onset sepsis.[7,8] In addition, many pathogens, particularly viral respiratory pathogens, have a distinct seasonality.[9] Because of this association between age and infection, we collect control specimens matching the distribution of young infants' ages when they become sick. Finally, for seasonal matching of cases, in each calendar month a certain number of enrolled newborns are selected for the control subcohort; this number is proportionate to the total number of live births in that month.

Operational Challenges With Control Selection

By design, all ANISA study data are collected using paper-based forms, which results in a significant lag between actual data collection and entry into the database. During study design, we were particularly concerned that newborns registered at age 0 or 2 days should have a specific probability of being selected as

Accepted for publication January 10, 2016.

From the *Department of Microbiology, Child Health Research Foundation, Dhaka, Bangladesh; †Centre for Child and Adolescent Health, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; and ‡CSL Software Resources Ltd., Dhaka, Bangladesh.

The ANISA study is funded by the Bill & Melinda Gates Foundation (Grant No. OPPGH5307). The authors have no other funding or conflicts of interest to disclose.

Address for correspondence: Shams El Arifeen, DrPH, Centre for Child and Adolescent Health, icddr,b, Mohakhali, Dhaka-1212, Bangladesh. E-mail: shams@icddrb.org.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.pidj.com).

ISSN: 0891-3668/16/3505-0S35
DOI: 10.1097/INF.0000000000001104

healthy controls at that age. This probability is conditional to the age distribution of pSBI cases and seasonal distribution of births. It would be difficult to carry this out without selection bias, given the delay in entering data from paper-based forms. Therefore, we opted for the mobile phone text message as a technological solution. This choice was driven by the simplicity of the technology, expected skills of the CHWs, and widespread availability and use of mobile phones in the targeted communities. The communities in the study sites have very few areas with no mobile phone coverage. An additional advantage of using text messages is that the technology works exactly the same way in all 5 ANISA study sites in 3 different countries and is compatible with all mobile phones.

CRITICAL STEPS FOR EFFECTIVE USE OF TEXT MESSAGES

Selecting healthy controls using text messages involves 3 steps: sending a text message for registration of a newborn in the database, running an algorithm based on selected data and generating a feedback text message to the CHW with the control selection status of the newborn (control or not control). The procedure is described in Figure 1. For these 3 steps to work and all data to link correctly at the time the newborn registration text message is sent, the database must receive the pregnancy registration information before sending the newborn registration text message. This is a challenge for ANISA because there is a delay in entry of the paper-based data capture forms (DCFs). Therefore, a CHW must also send a text message when enrolling a pregnant woman. This text message registers the pregnancy in the database in real time and creates a record that is linked to the newborn registration ID. If any newborn registration ID does not correlate with the corresponding pregnancy ID, the server does not accept the newborn ID, and the data are not processed. The second step involves the need to frequency match the age of the controls for providing specimens for the age distribution of pSBI cases with successful collection of blood and/or NP–OP specimens. We use the age distribution of pSBI cases as of the week before (to account for seasonality) the newborn is enrolled. Thus, upon enrollment and successful collection of specimens from a pSBI case, the study physician also sends a text message to the database server including the age of the young infant at the time of assessment. The control selection process is also self-adjustable with actual healthy control enrollment rates, particularly when there are shortfalls from the previous month.

Sending Text Messages to the Data Server

To address the 3 required steps mentioned earlier, a CHW sends a text message to a site-specific phone number at the time of registration of a pregnant woman or a newborn in the ANISA study. These messages include the mother's ID and the newborn's age. The study physician also sends a text message with the age of the newborn to the same phone number when he/she is successfully enrolled as a pSBI case or as a healthy control. Table 1 shows the different text message formats used in this system. A General

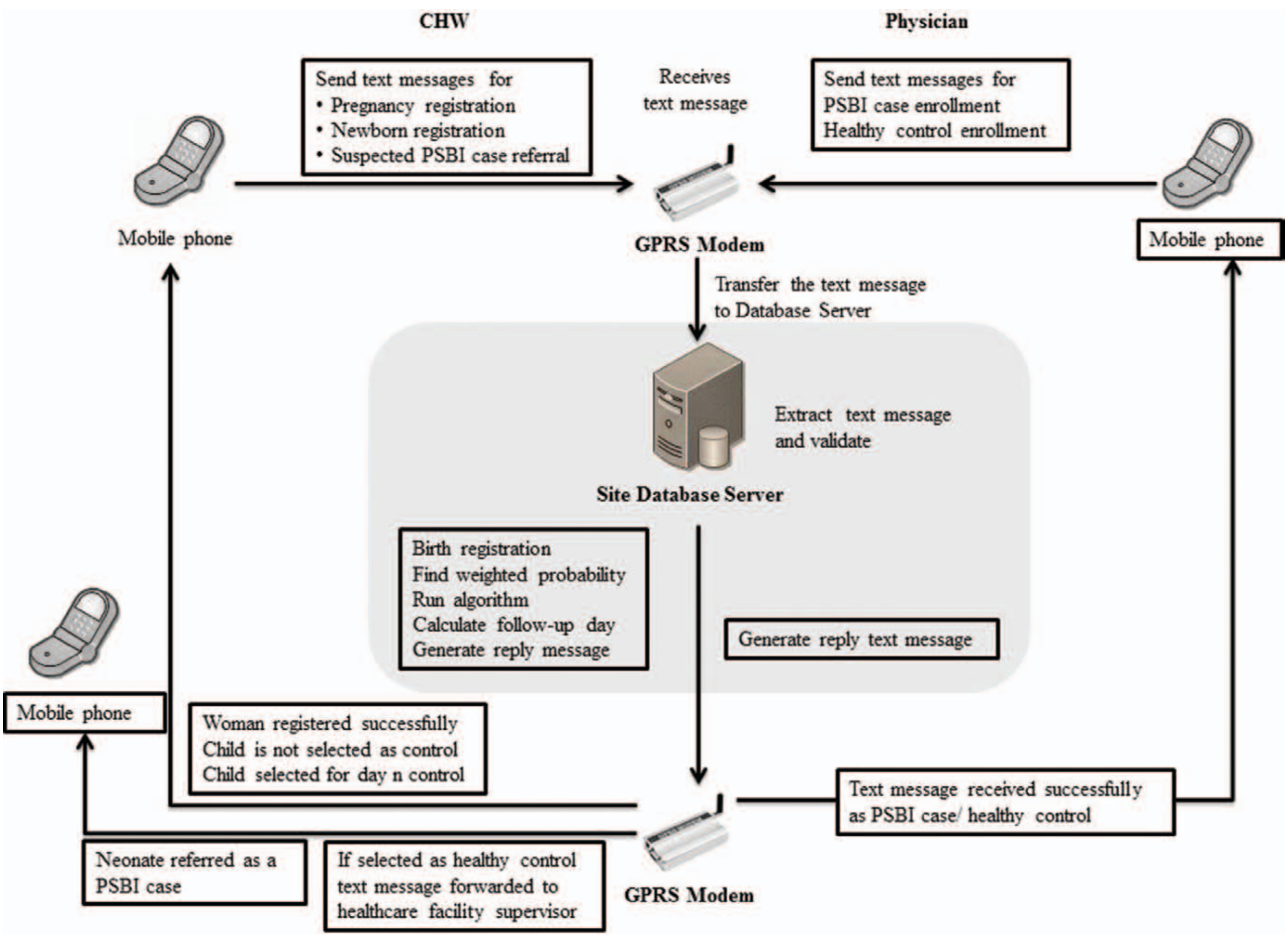

FIGURE 1. Architecture of text message system and control selection process in ANISA. full color online

TABLE 1. Text Message Formats for Updating the Database on ANISA Surveillance Activities

Type of Text Message	Text Message Formats	Sender
Pregnancy registration	P Middle five digit of Study ID	CHW
Newborn registration	C Middle five digit of Study ID # Child serial # Child Age	CHW
Suspected PSBI case referral by CHW	R Middle five digit of Study ID # Child serial	CHW
Healthy control enrollment (with specimen)	H Middle five digit of Study ID # Child serial # Child Age	Physician
PSBI case enrollment (with specimen)	S Middle five digit of Study ID # Child serial # Child Age	Physician

Packet Radio Service (known as GPRS) modem is used for text message transmission and data transfer to the site's server.

RANDOMIZATION OF NEWBORNS

Using an algorithm developed by the ANISA Data Coordination Center, a site's server randomizes each newborn at the time of registration into one of 2 subcohorts[10]: the control subcohort where the selected newborn will be asked to provide a NP–OP specimen or the noncontrol subcohort where the selected newborn will not need to provide clinical specimens unless he/she becomes sick. If a newborn is selected into the control subcohort, the algorithm determines at which age the newborn is required to provide specimens. This age will always coincide with the CHW home visits scheduled for that newborn. Once the randomization and date assignment process are complete, the server generates a return text message to inform the CHW to which subcohort that particular newborn will belong; and if selected for the control subcohort, at what age the selected newborn should provide specimens.

Principles of the Algorithm Used for Randomization

We decided to collect 400 control specimens (NP–OP) from each site. We estimated that the total number of registered newborns over a period of 2 years in each of the 5 study sites would range between 6000 and 20000. Therefore, between 2% and 7% of the registered newborns would need to provide specimens as healthy controls (referred to as the site-specific base probability). This probability is first equally divided (by 10) across all 10 visit ages. The (visit) age-specific base probability is adjusted every week according to the age distribution of enrolled cases up to the previous week. Because the actual ages of case enrollment differ from the visit ages, we created non-overlapping time frames around each scheduled visit age with no gaps between consecutive time frames. At this point in the process, the algorithm ensures that controls and cases are frequency-matched by age and season. The sum of all adjusted visit age-specific base probability adds up exactly to the overall base probability for the site (between 4% and 7%). This probability is adjusted based on the failure rate in the control specimen collection and any shortfall in actual cumulative enrollment up to the preceding week. A final adjustment (+3% additional selection) is made to allow for the controls that are eventually excluded by becoming ill within 7 days of control specimen collection. The number of births in most South Asian populations shows strong seasonality, usually peaking in late autumn and early winter. To adjust site-specific base selection probability for each week, we used historical records of seasonal variation in the number of births from each site to estimate the number of births expected and controls to be selected in each calendar week.

Step-by-step details of the algorithm are presented in Supplemental Digital Content 1, http://links.lww.com/INF/C405.

PROBLEMS AND SOLUTIONS

We identified several problems with using text messages in ANISA surveillance. The number of selected controls and their proportionality to the age distribution of cases may be jeopardized if study physicians fail to send text messages consistently. As indicated earlier, this input is 1 of the 2 critical steps for the control selection process. In addition, the parents of a significant proportion of selected young infants refuse to provide control specimens on the particular day for which they are selected by the system. The actual rate of refusal in enrolling controls (75%), is considerably higher than the assumption (30%) when designing the study. We increased the base control selection probability in 2 stages to account for the control enrollment failure. Another issue is that the data server calculates the age of an enrolled newborn when it receives a text message from a CHW. If there is a problem with phone network connectivity, the CHW will not receive any reply text message. To address this issue, we created an option in the ANISA data entry system to allow direct manual input of text messages by data management staff after CHWs provide the relevant information over the phone.

EXTENDED USE OF TEXT MESSAGES

The 3 steps of the control selection process prompt several additional uses of text messages, especially in real-time monitoring of critical study performance indicators. The real-time data generated by the text messages enable study coordinators and supervisors to keep track of pregnancies and newborns being registered and cases and controls being enrolled in the study. The data are disaggregated by a supervisor and the CHW, enabling identification of staff in need of additional support, monitoring and training. The data from text messages help monitor the data entry lag and locate missing forms. Initially, the average delay in entry of the newborn registration DCFs was over 60 days, which has decreased to fewer than 40 days with continuous monitoring through the text message data uploaded to the server (Fig. 2). Figure 3 shows the numbers of pregnancies recorded in the database from text messages and paper-based DCFs at 4 different time points. This chart is characterized by dips in

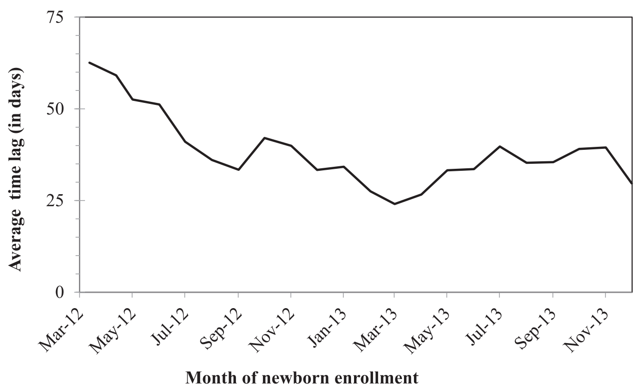

FIGURE 2. Monthly mean time lag between registration of birth with text messages and data entry from paper-based DCFs, March 2012–November 2013.

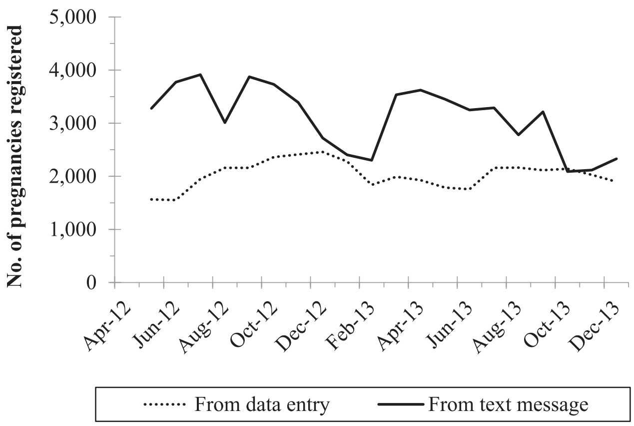

FIGURE 3. Number of pregnancies recorded in the database from text messages and from paper-based form entry, May 2012 –December 2013 (Bangladesh and Pakistan sites).

the 2 or 3 months before each peak. Also, in the first month (June 2012), the low pregnancy registration rate indicates that the study site data teams were still learning to manage the processing of the paper-based DCFs. The text message system helps in reducing the workload of project staff, as study physicians receive alert messages before referral, so that they do not have to be available on-site around the clock. This is particularly helpful during holidays. Similarly, text messages alert the laboratory staff about the arrival of specimens. As ANISA study sites include remote areas and specimens are collected around the clock, laboratory staff need to be informed in advance about incoming specimens and tentative times of arrival so that they can prepare for specimen processing. Finally, using the text message data, study supervisors can track young infants who do not comply with referral and arrange follow-up visits.

SUMMARY

Use of text messages for health purposes has mostly occurred in developed countries while the experience from low-income countries remains mostly anecdotal although some recent studies have used text message systems to collect data from communities. In ANISA, we are using a text message system in a large cohort for real-time data entry and monitoring study progress in resource-poor settings. The methodology described in this paper allows us to randomize the newborns enrolled in the study based on age and seasonal distribution of cases. In addition, this system is very useful in off-site study supervision and monitoring. We believe that extended use of this mobile phone technology in community-based public health research can improve study performance and data quality.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Bill & Melinda Gates Foundation as well as the technical assistance of our colleagues at the Child Health Research Foundation, Dhaka, the International Centre for Diarrhoeal Disease Research, Bangladesh and the Centers for Disease Control and Prevention.

REFERENCES

1. Jansen RR, Wieringa J, Koekkoek SM, et al. Frequent detection of respiratory viruses without symptoms: toward defining clinically relevant cutoff values. *J Clin Microbiol*. 2011;49:2631–2636.
2. Bosch AA, Biesbroek G, Trzcinski K, et al. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog*. 2013;9:e1003057.
3. Mathisen M, Strand TA, Valentiner-Branth P, et al. Respiratory viruses in nepalese children with and without pneumonia: a case-control study. *Pediatr Infect Dis J*. 2010;29:731–735.
4. Bogaert D, van Belkum A, Sluijter M, et al. Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. *Lancet*. 2004;363:1871–1872.
5. Robinson J. Colonization and infection of the respiratory tract: What do we know? *Paediatr Child Health*. 2004;9:21–24.
6. Islam MS, Baqui AH, Zaidi AK, et al. Infection surveillance protocol for a multicountry population-based study in South Asia to determine the incidence, etiology, and risk factors for infections among young infants 0 to 59 days old. *Pediatr Infect Dis J*. 2016;35 (Suppl 1):S9–S15.
7. Stoll BJ, Hansen NI, Sánchez PJ, et al.; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Early onset neonatal sepsis: the burden of group B Streptococcal and *E. coli* disease continues. *Pediatrics*. 2011;127:817–826.
8. Bizzarro MJ, Raskind C, Baltimore RS, et al. Seventy-five years of neonatal sepsis at Yale: 1928-2003. *Pediatrics*. 2005;116:595–602.
9. Gardinassi LG, Marques Simas PV, Salomão JB, et al. Seasonality of viral respiratory infections in southeast of Brazil: the influence of temperature and air humidity. *Braz J Microbiol*. 2012;43:98–108.
10. Rahman QS, Islam MS, Hossain B, et al. Centralized data management in a multicountry, multisite population-based study. *Pediatr Infect Dis J*. 2016;35 (Suppl 1):S23–S28.