

RESEARCH ARTICLE

Open Access



Use of the data system for field management of a clinical study conducted in Kolkata, India

Ju Yeon Park¹, Deok Ryun Kim^{1*}, Bisakha Haldar², Aiyel Haque Mallick², Soon Ae Kim¹, Ayan Dey¹, Ranjan Kumar Nandy², Dilip Kumar Paul³, Saugata Choudhury³, Shushama Sahoo³, Thomas F. Wierzbza^{1,4}, Dipika Sur², Suman Kanungo², Mohammad Ali^{1,5} and Byomkesh Manna²

Abstract

Background: Designing an appropriate data system is important to the success of a clinical study. However, little information is available on this topic. We share our experiences on designing, developing, and implementation of a data system for management of data and field activities of a complex clinical study.

Methods: The data system was implemented aiming at determining the biological basis for the underperformance of oral vaccines, such as polio and rotavirus vaccines in children at a site in Kolkata, India. The system included several functionalities to control data and field activities. It was restricted to authorized users based on their access privileges. A relational database platform was chosen, and Microsoft Visual FoxPro 7.0 (Microsoft Corporation, Seattle, WA, USA) was used to develop the system. The system was installed at the clinic and data office to facilitate both the field and data management activities.

Results: Data were doubly entered by two different data operators to identify keypunching errors in the data. Outliers, duplication, inconsistencies, missing entries, and linkage were also checked. Every modification and users log-in/log-out information was auto-recorded in an audit trail. The system offered tools for preparation of visit schedule of the participants. A visit considered as protocol deviation was documented by the system. The system alerted field staff to every upcoming visit date to organize the field activities and to inform participants which day to come. The system also produced a growth chart for evaluating nutritional status and referring the child to a specialized clinic if found to be severely malnourished.

Conclusion: The data system offered unique features for controlling for both data and field activities, which led to minimize drop-out rates as well as protocol deviations. Such system is warranted for a successful clinical study.

Keywords: Data management, Field management, Database, Data system, Clinical study

Background

Clinical studies play important role for improving human health; therefore, these studies should be managed properly and carefully [1, 2]. Data collection, data cleaning, editing, and management of the data in compliance with regulatory standard and the International Conference on Harmonization Harmonized Tripartite Guideline for

Good Clinical Practice (ICH GCP) are important aspects of a clinical study. A data system should ensure that these issues are well addressed in the system design [3]. The design should also ensure that the data are accurate, complete and compliant with regulatory standards and GCP requirements, and the analyses are done using cleaned data sets. Besides, the system should include tools for providing support to the clinic, field and laboratory staff in order to collect data accurately and in a timely manner. Only a few trials goes exactly as initially planned. For instance, a case report form (CRF) may need updating in

*Correspondence: drkim@ivi.int

¹ International Vaccine Institute, Seoul, South Korea

Full list of author information is available at the end of the article

the course of trial, a new trial site may be added, and new technology may emerge [4]. An investigator is, therefore, to worry about installing a data management system that is flexible and compliance with set guidelines and standards.

Although literature is plentiful in describing various aspects of data analysis, little can be found that tells about practical aspects of designing a data system [5–8]. Rarely, a data system includes tools to perform both data and field activities. It is worth mentioning, analyses of the clinical studies can be flawed not only by problems in data acquisition and field methodology, but also by errors in the design of the data system.

Recent hardware and advanced software tools have made it possible to develop an ideal data system for clinical studies [5]. In designing the system, one should consider study objectives, nature of the study site, and the local issues. The design should be flexible to accommodate unanticipated issues in the field and data management. Inclusions of tools for checking data outliers and inconsistencies as well as for queries are essential. Storing other information such as metadata (data dictionary) and audit trail, that tracks all the modifications ever made in the system, must be included in the system. While designing, one should also focus on system specifications, structure of the database, data and field management tools, and performance of the system.

We designed, developed, and implemented a data system for a complex clinical study in Kolkata, India. In the study, the vaccines was given to the participants concomitantly with the EPI routine immunization program for a period of 1 year, and there were multiple visits for specimen collection from the study participants, which could have been well managed with the support from the data system. This paper discusses the design, development, and implementation, as well the performance of the data system.

Methods

The clinical study

The clinical study named provide (Performance of Rotavirus and Oral Polio Vaccine in Developing Countries) was initiated in Kolkata, India with the hypothesis that there are factors for the underperformance of oral vaccines in children living in the developing world. It aimed to determine whether decreased vaccine responsiveness to oral poliovirus or rotavirus vaccines is associated with the presence of tropical enteropathy (TE); and to evaluate whether the impact of an inactivated polio vaccine (IPV) boosted systemic and mucosal immune responses to polio vaccines following vaccination with oral polio vaccine (OPV) in children with and without TE. The study was carried out among infants residing in Kolkata.

The study aim was to recruit 372 children at 6 weeks of age and follow them until they were 53–54 weeks of age. There were 12 visits for each of the participants during the follow-up time including the initial visit. This study was also carried out in accordance with the routine immunization schedule in India. Therefore the participants would have to be vaccinated with other vaccines, such as, BCG, DPT, HepB and Measles.

The data system

A custom-made data system, named INDSys, was designed and developed aiming to turn the information from the participants into data, to transcribe the data into a database without errors, and to manage data delivery from clinics and laboratory efficiently. The design phase focused on defining the database components and the modules and interfaces required for satisfying the need of data and field activities. While developing the system, provisions were kept for accommodating any new data or field related issues into the system. The system was developed using Microsoft Visual FoxPro 7.0 (Microsoft Corporation, Seattle, WA, USA), and the data was managed in a relational database environment. All related data, including clinical, laboratory and other data sets as well as field work schedule, were integrated within the data system.

The data system was restricted to authorized users only such as data managers, data operators, clinical monitors and investigators. There were two levels of restrictions in the system. At the first level, the users required login identification and password to access the system. At the 2nd level, the system functionalities were restricted according to the user's privileges. The functionalities according to the user's role are shown in Table 1. The data entry screens were designed to look similar to the data forms for facilitating the data entry operations. A dual data entry system was designed to avoid keypunching errors in the data [9]. A comprehensive data validation tool was incorporated in order to identify missing values, outliers, duplications, and intra- and inter-record inconsistencies in the data. Errors that cannot be corrected, such as unordinary data or missing laboratory results in the system, were also selected to remain in the database with proper documentation. The data system automatically stored information on changes in the data, users' log-in, log-out, entry, modification, and deletion of the data in an audit trail. Information such as who changed the data and when the changes were made were also stored in the audit trail. Additional information such as a data dictionary, data status, data checking plan and code plans for open-ended questions were accessible through the system. A report generation module was included in the system to monitor the performance of field and data activities.

Table 1 Accessible functionalities according to the user's role

Role	Main responsibilities	Accessible functionalities
Database administrator	Grant permission to a user for accessing system and maintain system and troubleshooting	All functions
Data manager	Supervise data entry, error check, and backup	Entry, view/edit/delete, dual check, error check, report/query, maintenance(backup, change password)
Data operators	Data entry and check dual entry error	Entry, view/edit, dual check, maintenance (change password)
Clinical monitor(s)	View data for monitoring electronic data	View, report/query, maintenance (change password)
Investigator(s)	Supervise all data activities	View, report/query, maintenance (change password)

For the field management, the system design included activities in field office, such as participant's contact, calendar of all visits, participant's discontinuation/drop-outs, growth chart, missing visits, etc. Necessary tools (*vide infra*) were included in the system so that the field management could control the activities easily and smoothly, especially for this complicated schedule of visits. Each participant was to be vaccinated with several vaccines (e.g. DPT, HepB, and Oral Polio Vaccine) in accordance with Expanded Program on Immunization (EPI) in India, and study vaccines (e.g., Rotavirus, Oral Polio Vaccine, or Inactivated Polio Vaccine). If a participant received a vaccine on a day other than the scheduled day, the subsequent date of vaccination as well as scheduled dates of visit for specimen collection would also change to the date of vaccination and a new calendar for the visitation schedule was prepared for the participant accordingly. The data system included tools to generate the new visitation schedule after each visit was made by a participant.

Implementation of the data system

The data system was implemented in standalone computers at the data center and at the field clinic of the National Institute for Cholera and Enteric Disease (NICED). Microsoft Windows was chosen as the operating system to implement the data system. No other software was required to operate the system. The data and field staff of the project were trained on how to operate the data system. Authorized users, had access to the system through assigned identification and password. The staff members at the clinic had access to the system only for management of the field activities. The database at the clinic was continuously updated by the data staff of the project to facilitate field activities with updated data sets.

Ethical considerations

Written informed consent was obtained by mothers of participating children. The study protocol was approved by the scientific advisory committee (SAC), Institutional Ethics Committee of the National Institute

of Cholera and Enteric Diseases (NICED), the Health Ministry Screening Committee of India, and the International Vaccine Institute Institutional Review Board. The protocol was registered at clinical trial registry of India (CTRI/2012/03/002504) and at clinicaltrials.gov (NCT01571505).

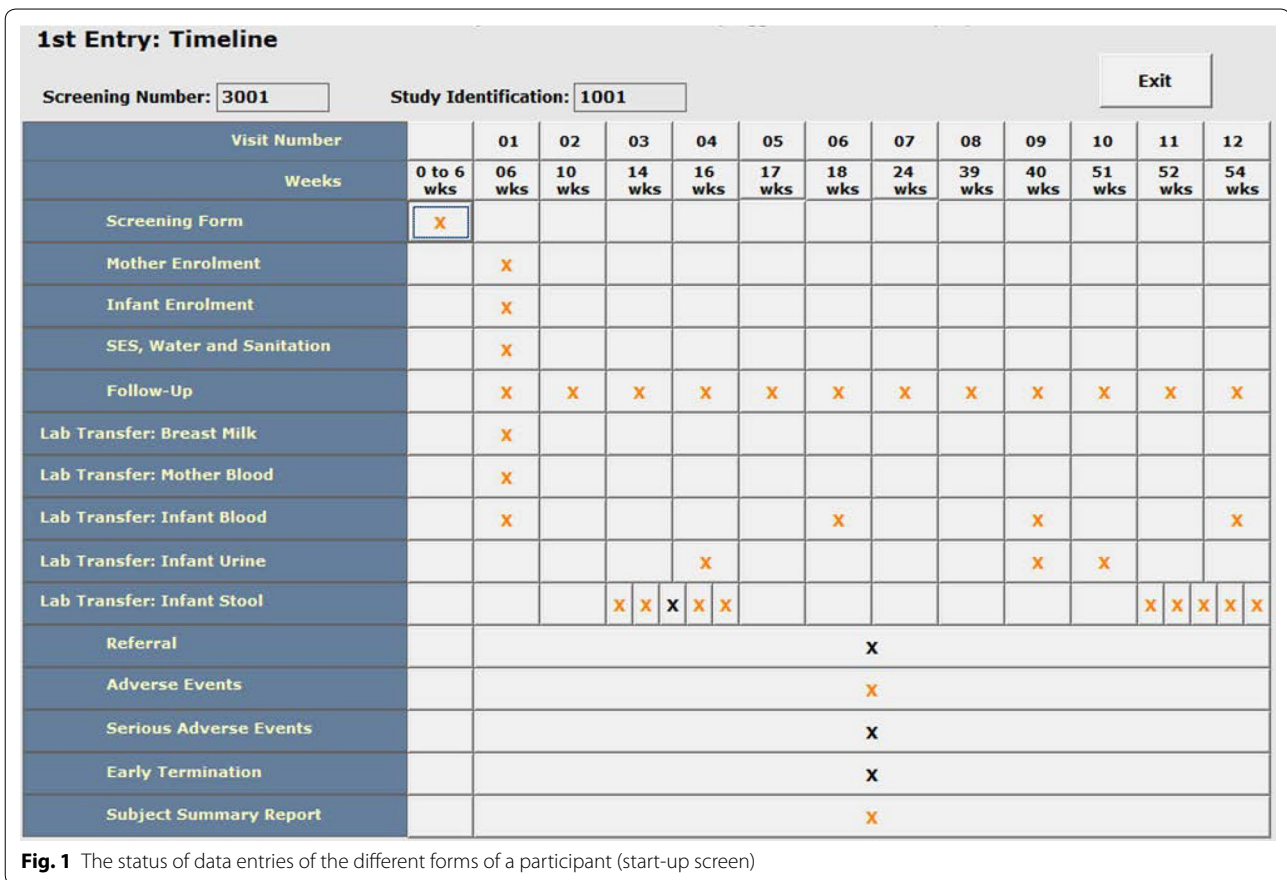
Results

Start-up screen

Upon starting the system, a log-in screen appears for accessing the system. After successful log-in, a menu describing the list of activities appears on the screen according to the user's privileges. As a dual data entry system, the system provides options for entering the data either in the 1st or 2nd file according to the user's privileges. If the operator enters a Screening Number for a new recruitment, the system shows the Data Entry Status screen (called start-up screen) of the participant (Fig. 1). In this screen, the data entry status for the different visits of the participant is reflected by color, and the number of successful and missed visits are easily identifiable. Since the screening form is used to assign the Study ID, the user will receive an error message if s/he tries to enter any other forms before entering the screening form. In the event of early termination of a participant, the operators are not allowed to enter any other forms of the participant. When one-year of follow-up of a participant is passed, the system checks for missing visits and specimens, and alerts accordingly. This helps ensure completeness of the data entries of all visits of a participant.

Data entry and editing

The study had two sources of data: clinic and lab. The data from the clinic, such as enrolment, follow-up, and specimen collection were entered into the data system soon after receiving the forms. The lab data were entered and verified by lab scientists following data validation rules and then uploaded onto the system. The embedded data validation process in the system ensures quality of the data after checking for outliers and inconsistencies.



While entering the data from the clinic, the other functions are inactivated by the system. In the event of a potential error during data entry, the system warns the operator and provides options for correcting it. However, the operator can move on without correcting the data if the correction cannot be made at his/her level. A specific module for data checking is incorporated into the system (vide supra) to trap any error in the database.

The system provides a view screen for reviewing all data forms in one screen. The user can view the data by using navigation keys located on the right side of the page of the screen and selecting the participant ID, or by entering the participant ID in a specified box. If a participant ID is selected, the participant's data form can be viewed by clicking on a specific tab of the form abbreviations.

The lab data, before final submission to data room, were recorded on specific data forms designed for lab results, dually entered in Excel, and further checked by lab personals. The data generated by the laboratory is uploaded onto the system from the soft copy of the data. In this case, the user clicks onto the Upload Lab Result tool to get the list of the file names of all the lab data. If a wrong file name is selected, the system warns the user and prevents him/her from uploading the data.

Unlike clinic data, the lab data are not allowed to edit directly on the screen. If an error is detected in the lab data, the error is sent to the laboratory for correction. After getting the corrected data in a soft copy, the erroneous data was replaced with the corrected one.

Data validation and audit trail

There were two modules in the system for checking accuracy of the data. Firstly, the data are doubly entered to detect keypunching errors. The system provides a list of discrepancies between the two entries including unmatched unique identification of the children between the two entries. Once the keypunching errors in the data are resolved, the user is allowed to move to the second step of data checking which includes outliers, inconsistencies, validity of the dates, and linkage. If the errors generated at the source, the forms along with the type of errors are sent to the field office for resolution. Once the error feedbacks are received, the data are updated accordingly. When an update is done, the user ID, date and time of update, and old and new values are stored in the audit trail. Audit trails are being created incrementally, in chronological order, and in a manner that it does not allow new audit trail information to overwrite

existing data. If updates are made multiple times, the users can trace complete history of the updates.

Producing schedule of visits

As mentioned above, there were 12 visits for each participant in a year including the initial visit for consent and screening of the participants. It was necessary to calculate the scheduled dates of visit for each participant during the follow-up period, and remind them to come to the clinic on that date. Since enrolment dates differed from one participant to another and the purpose of visit also differed from one visit to another, keeping track of the schedule dates and purpose of visit would be difficult without having field management tools in the system. Note that the scheduled dates of visit might also change depending on the last date of visit. The system produces a fresh calendar of the scheduled dates of appointments after every visit is made (Fig. 2). This helped Field Supervisors to fill the dates of visit schedule in the Study Identification Card which should be brought when the participants come to the clinic, to remind them by telephone and home visit prior to the schedule date, and to identify whether the visits were made on the planned date or not. The system also produces the list of all visits

to be made on the upcoming clinic day (an example is shown in Fig. 3), so that the field staff can take necessary actions to make those visits successful.

Data backup

Data can be lost by a computer accident or catastrophic loss -such as losing a critical file or experiencing unexplained file corruption, a hard drive crash, or total system loss. The system provides a backup tool to maintain a regular backup at the end of each working day. A backup copy was kept in a separate building away from the data center.

Reporting protocol deviation

One of the most challenging tasks of the clinical study was to manage the scheduled dates of visits of the participants for vaccination and specimen collection. The target visit dates may change depending on the previous date of visit of the participants. For visit week 6–18, participants will be invited to come to the clinic based on the date of his/her last vaccination on account of the mandatory 4 weeks window between vaccinations. For visit week 24–53, participants will be at age of the visit week +7 days, regardless of the date of last vaccination.

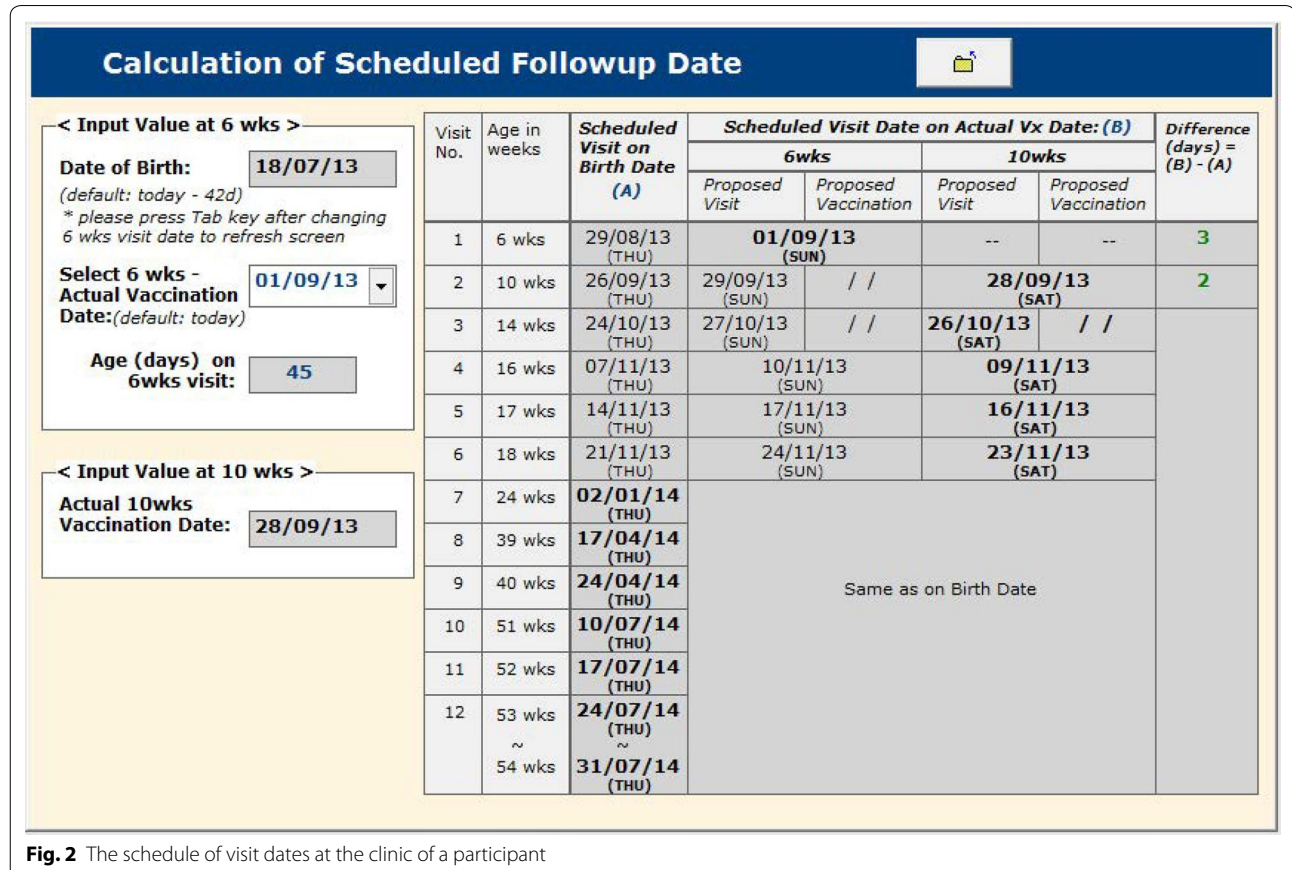


Fig. 2 The schedule of visit dates at the clinic of a participant

Followup Schedule																							
IVI/PROVIDE STUDY/NICED, INDIA																							
Proposed Visit / Vaccination Date=26/08/13 (Monday)																							
SID	Clinic Visit Week	Actual Vx date						Proposed Date		Anthropometry		Vaccination						Specimen Collection					
		6wks (OPV)	10wks (OPV)	14wks (CPV)	17wks (Rota)	39wks (OPV/1PV)	52wks (CPV)	Lower bound	Upper bound	Weight	Length	DTP	HepB	OPV	IPV	Rota	Measles	VitA	Mother		Infant		
		BM	Blood	Blood	Urine	Stool																	
1091	53-54 wks	28/09/12	31/10/12	30/11/12	20/12/12	17/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1092	53-54 wks	28/09/12	31/10/12	30/11/12	20/12/12	17/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1093	53-54 wks	28/09/12	31/10/12	30/11/12	20/12/12	17/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1094	53-54 wks	28/09/12	31/10/12	03/12/12	19/12/12	17/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1096	53-54 wks	01/10/12	31/10/12	30/11/12	20/12/12	23/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1097	53-54 wks	01/10/12	31/10/12	03/12/12	21/12/12	23/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1098	53-54 wks	01/10/12	02/11/12	03/12/12	21/12/12	23/05/13	19/08/13	25/08/13	27/08/13	Y	Y									Y			
1103	52 wks/Day 11	03/10/12	02/11/12	05/12/12	26/12/12	16/05/13	//	25/08/13	27/08/13													Y	
1106	52 wks/Day 4	05/10/12	07/11/12	07/12/12	27/12/12	23/05/13	//	25/08/13	27/08/13													Y	
1112	52 wks/Day 0	10/10/12	09/11/12	10/12/12	02/01/13	30/05/13	//	23/08/13	26/08/13	Y	Y			Y								Y	
1113	52 wks/Day 0	10/10/12	09/11/12	10/12/12	02/01/13	30/05/13	//	23/08/13	26/08/13	Y	Y			Y								Y	
1115	51 wks	17/10/12	16/11/12	17/12/12	05/01/13	06/06/13	//	26/08/13	02/09/13	Y	Y											Y	
1200	40 wks	02/01/13	01/02/13	04/03/13	28/03/13	//	//	25/08/13	27/08/13	Y	Y					Y	Y			Y	Y		
1268	24 wks	22/04/13	24/05/13	24/06/13	16/07/13	//	//	26/08/13	02/09/13	Y	Y												
1294	18 wks	30/05/13	01/07/13	31/07/13	//	//	//	25/08/13	27/08/13	Y	Y											Y	
1301	17 wks	07/06/13	08/07/13	07/08/13	//	//	//	26/08/13	02/09/13	Y	Y				Y								
1310	14 wks/Day 0	26/06/13	26/07/13	//	//	//	//	23/08/13	26/08/13	Y	Y	Y	Y	Y								Y	
1332	10 wks	26/07/13	//	//	//	//	//	26/08/13	02/09/13	Y	Y	Y	Y	Y	Y								
1333	10 wks	26/07/13	//	//	//	//	//	26/08/13	02/09/13	Y	Y	Y	Y	Y	Y								

Fig. 3 List of the participants to be visited on the upcoming clinic day

If a visit or procedure cannot be made within the specific time period (called window period), then those visits or procedures are treated as protocol deviation. Since this clinical study depends largely on the biological issues, any protocol deviation would have negative impact on the results of the analysis. Therefore, it was important to keep the number of protocol deviations as minimized as possible.

The data system produces how many visits or procedures turned out to be protocol deviation. Note that the system generated protocol deviation is data driven, thus the study could overcome limitation of manually created protocol deviation form, which is prone to error and may differ from the data actually stored in the database. An example of the system generated protocol deviation is shown in Fig. 4. The system also provides the status of all visits, vaccinations, and sample collections with (in red color) and without (in green color) protocol deviation (Fig. 5).

Reporting of severely malnourished children

The study required referring severely malnourished children to a specialized clinic for treatment. Upon request, the system provides an anthropometric data calculator for calculating nutritional status and drawing growth charts of the children, which were evaluated

to understand the nutritional status of the children. The system produces two types of growth chart for each participant: weight-for-age and length-for-age. By triggering the participant’s ID, the user can instantly get the child’s growth status according to the World Health Organization (WHO) child growth standard (<http://www.who.int/childgrowth/standards>), as shown in Fig. 6. In the figure, the two lines from the bottom were related to malnutrition status. If the child’s weight or height is marked below the line of SD2neg (<-2SD), the child is considered malnourished. If the child’s weight or height is marked between line of SD2neg and SD3neg, the child is considered moderate malnourished. And, if the child’s weight or height is marked below SD3neg (<-3SD), the child is considered severely malnourished. In case of severely malnourished, the child was referred to a specialized clinic for treatment. However, with this growth chart, the user could not get z-scores for anthropometric indicators (weight-for-age and height-for-age) based on the WHO child growth standard. Thus, the list of severely malnourished children generated from the system was just to facilitate field activities (Figs. 7, 8).

Discussion

Running a clinical study smoothly like ours is challenging due to its complex and long duration visit schedule.

Form: KV_PDV
Version 1.9; February 1, 2011

SID: KVC1065

Underperformance of Oral Polio and Rotavirus Vaccines Study, NICED, India

Protocol Deviation Case Report Form

1	Date of Report	23/08/13	
2	Form completed for	1 = Mother 2 = Infant	<u>2</u>
3	Date of Protocol Deviation	26/07/13	
4	Visit week associated with protocol deviation	01=Enrollment 02=Week 06 03=Week 10 04=Week 14 05=Week 16 06=Week 17 07=Week 18 08=Week 24 09=Week 39 10=Week 40 11=Week 51 12=Week 52 13=Week 53-54 99=Not applicable	<u>12</u>
5	Type of deviation?	01=Missed visit 02=Missed specimen 03=Missed shedding specimen 04=Missed procedure 05=Missed specimen + procedure 06=Blood specimen collected outside visit 07=Urine specimen collected outside visit 08=Stool specimen collected outside visit window 09=Breast milk specimen collected outside visit window 10=Visit occurred outside of window 11=Vaccination outside clinic 12=Incorrectly vaccinated 13=Child received additional dose of OPV 14=Other	<u>02</u>
	5a. If any, please specify	02=Missed specimen: Infant Stool/Day 25	
6	Reason for deviation?	1 = Subject illness 2 = Subject unable 3 = Subject refusal 4 = Clerical 5 = Investigator decision 6 = Other	<u>6</u>
	6a. If any, please specify		
7	Did the deviation result in an adverse event?	1 = Yes 2 = No	<u>2</u>

Fig. 4 A protocol deviation form generated from the data system

In our study, all 372 participants were successfully recruited according to the planned schedule. The drop-out rates were only 8 % and protocol deviations stood at 3 %. The reasons for these protocol deviations included:

visits occurring outside the window period (44 %), missed specimen (34 %), missed visits (11 %), specimen collected outside the window period (10 %), and missed procedures (1 %). 99 % blood samples, 95 % stool samples, 98 %

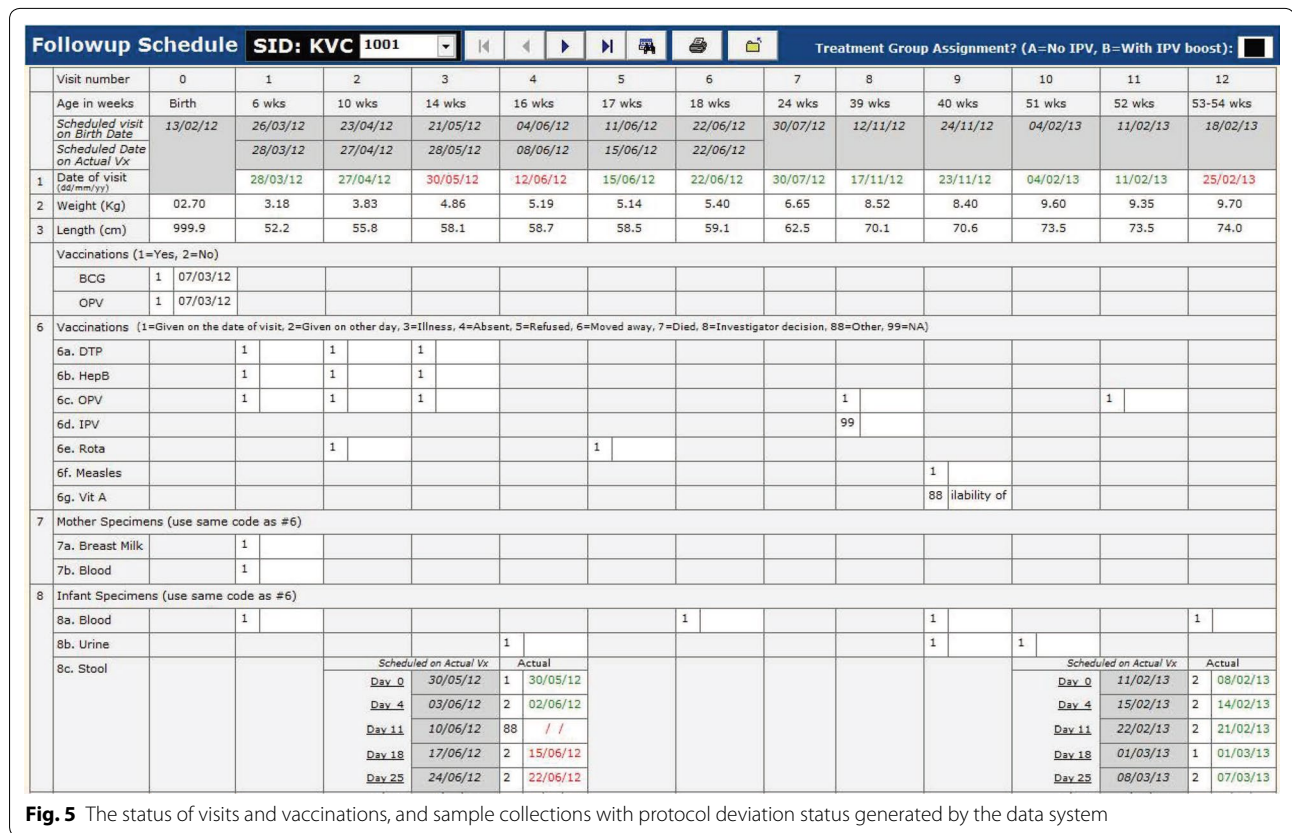


Fig. 5 The status of visits and vaccinations, and sample collections with protocol deviation status generated by the data system

urine samples, and 100 % breast milk samples were successfully collected. Such an excellent performance would have not been possible without having the support provided by the data system.

In addition to manage data activities and generate data reports, our data system provides support for management of the field activities such as preparing visit schedule, alerting next day visit schedule and assessing requirement of the logistics including number of doses of the vaccines, etc. for the next visit schedule. It also ensures a good clinical practice, a legal requirement to conduct clinical trials in many countries. The other advantage in our system is that it generates nutritional status report of the participants at each time point so as to take immediate actions against severely malnourished infants, such as referring to a specialized clinic where proper treatment can be ensured. In the system, the data interoperabilities are limited to Acrobat Reader and Excel, as required by us. However, we have kept provision for the data interoperability in any other standard data format.

Our data system is specific to our study, which is unlikely to match exactly with any other studies. Three-month time of a skilled system designer and programmer was required to design and develop the system. Unlike

open source software where many developers have the opportunity to scrutinize the system, our system was scrutinized in-house.

Since many developers scrutinized the open source application, it is hard for bugs to hide in that application. Also, since codes of the open source software are shared among numerous parties, it is typically well structured, which cannot be ensured a single company develops like us [10].

However, our system passes through different courses of action over the time, and we were able to fix the bugs in the system. More importantly, we were able to tailor it specific to the project activities, which may not have been possible in an open source software solution.

The limitation in our system is that it was installed on standalone computers. Therefore, we had to upload of the updated database to the field office computer on a regular basis. Similarly, the data generated at the lab were uploaded manually into the system. This offline mechanism could have been avoided if the computers at the clinic and lab were connected to the data office through a network. However, we could not make it possible in our setting due to limitation of resources.

Still a large part of clinical centers uses their own developed solution or a single solution [11] because

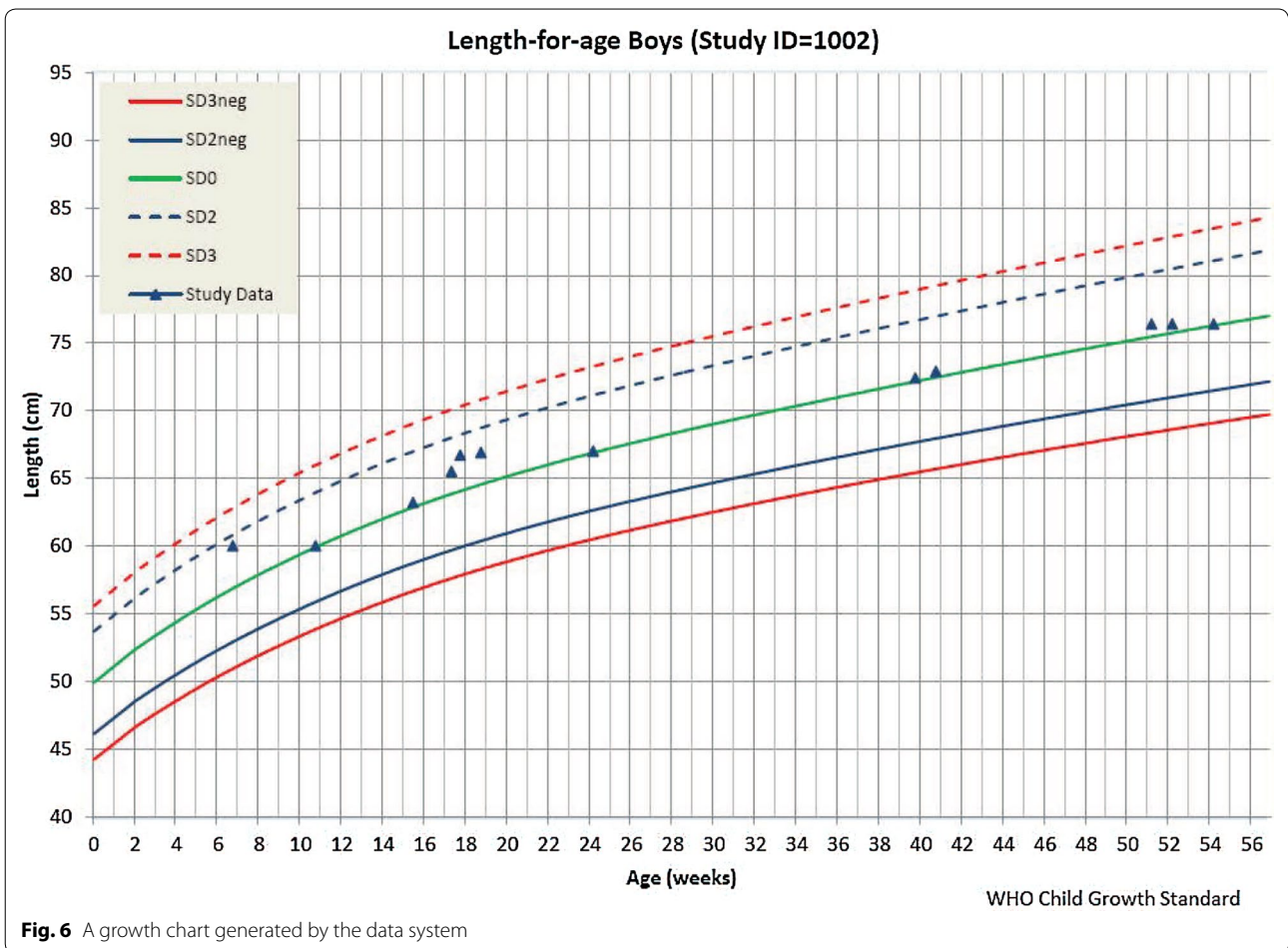


Fig. 6 A growth chart generated by the data system

Anthropometric Calculator

Enable to derive nutritional status information based on WHO standard for weight-for-age (WAZ), length/height-for-age (HAZ), and BMI-for-age, and print out the list of malnourished children (z-score is below -3SD)

<From Database>

Printing ANY of Z-Scores < -3SD:

Begin* (dd/mm/yyyy): 08/10/2012
 Default is the Monday of the week

End (dd/mm/yyyy): 22/10/2012
 Default is today

*Leave blank date field to print from First Enrolment Date to Today

<At the Clinic>

Calculating Z-Scores: please enter the information....

Date of birth: 27/02/2012 (default: Today) **Age: 6 weeks 2 days**

Visit Date: 11/04/2012 (default: Today) **Age: 44 days**

Gender: 1 = Male 2 = Female

Weight (kg): 3.25

Length (cm): 54.2

BMI: 11.06

Weight-for-age (WAZ): -3.19

Length-for-age (HAZ): -1.16

BMI-for-age: -3.78

Buttons: OK, Clear, Cancel

Fig. 7 An anthropometric calculator

List of any Z-Score <-3SD**IVI/PROVIDE STUDY/NICED, INDIA**

Study ID	Clinic Visit Week	Gender (1=M;2=F)	Age (days)	Weight (kg)	Length (cm)	BMI	Weight -for-age	Length -for-age	BMI -for-age	Referral Form Exists (1=Yes)	Comments
1152	39 wks	1	276	5.94	60.4	16	-3.614	-5.191	-0.653		
1152	40 wks	1	283	6.02	60.5	16	-3.564	-5.260	-0.508	1	
1312	06 wks	1	44	3.22	52.4	11	-3.246	-2.070	-3.206	1	
1328	06 wks	1	43	3.20	50.3	12	-3.230	-3.076	-2.364	1	
1336	06 wks	1	45	2.86	50.6	11	-4.023	-3.042	-3.722	1	

Fig. 8 The list of participants malnourished children (z-score is below -3SD) generated by the system

clinical data managements are very heterogeneous, and the open source solutions do not play a major role in clinical trial data management [12]. However, because of flexibility, increased innovation, shorter development times, and faster procurement processes, open source software may be attracted by an organization. Also by using open source software, an organization will not be locked into using a proprietary software program. One disadvantage in an open source solution is that, due to the web-based nature of this system, it may pose a challenge for real time data entry because in many developing country settings internet connectivity is a problem. One may also face sluggish response times and system timeouts. Additionally, XML rules may adversely affect the application's response time and therefore other options may need to be explored and used while using open software solution [3].

Conclusion

Clinical data management has evolved and will continue to do so in response to need [13]. Limited literature hinders the capacity of scientists to design and develop a well-managed data system for their studies [14]. An ill-managed data system may lead to false outcomes, which is detrimental for the study; eventually for human health. The concepts and ideas we discussed in this paper may be useful for designing and developing a well-managed data system for the clinical studies. By controlling both the data and field activities in a system, the investigators may overcome the complexities of the visit schedules in their studies. We believe such a system would be useful for the investigators who want to initiate a complex clinical study.

Availability of supporting data

This paper describes the design, development, and implementation of a data system. The source codes of the data system and operators' manual can be made available upon request.

Abbreviations

GCP: good clinical practice; TE: tropical enteropathy; IPV: inactivated polio vaccine; OPV: oral polio vaccine; EPI: expanded program on immunization; ID: identification; WHO: World Health Organization; SD: standard deviation.

Authors' contributions

DRK and MA contributed to the design of the data system. DRK, BM, and MA contributed to the implementation of the system. DRK, BM, SK, BS, AHM, SAK, AD, RKN, DKP, SC, SS, TFW, DS and MA supervised the study activities. JYP, DRK, and MA analyzed the system tools, reviewed the performance of the system, and wrote the first draft of the manuscript. All authors participated in the writing of the manuscript and had access to the data system. All authors read and approved the final manuscript.

Author details

¹ International Vaccine Institute, Seoul, South Korea. ² National Institute of Cholera and Enteric Diseases, Kolkata, India. ³ B.C. Roy Post Graduate Institute of Pediatric Sciences, Kolkata, India. ⁴ PATH, Washington, DC, USA. ⁵ Johns Hopkins Bloomberg School of Public Health, Baltimore, USA.

Acknowledgements

We are grateful to the people of the study area in Kolkata and our field staffs. The study was funded by the Bill & Melinda Gates Foundation. The Foundation had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 22 May 2015 Accepted: 30 November 2015

Published online: 09 January 2016

References

- Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS. Data management in clinical research: an overview. *Indian J Pharmacol.* 2012;44(2):168–72.
- La HO. Quality management system of clinical research. *Kor J Clin Pharmacol Ther.* 2006;14(1):23–8.
- Ngari MM, Waithira N, Chilengi R, Njuguna P, Lang T, Fegan G. Experience of using an open source clinical trials data management software system in Kenya. *BMC Res Notes.* 2014;7:845.
- Ene-Iordache B, Carminati S, Antiga L, Rubis N, Ruggenenti P, Remuzzi G, Remuzzi A. Developing regulatory-compliant electronic case report forms for clinical trials: experience with the demand trial. *J Am Med Inform Assoc.* 2009;16:404–8.
- Pradhan EK, Katz J, LeClerq SC, West KP Jr. Data management for large community trials in Nepal. *Control Clin Trials.* 1994;15(3):220–34.
- Youngblut JM, Loveland-Cherry CJ, Horan M. Data management issues in longitudinal research. *Nurs Res.* 1990;39(3):188–9.
- Goodger WJ, Bennett T, Garcia M, Clayton M, Pelletier J, Eisele C, Thomas C. Development of a database management/analysis system for field

- research activities within a coordinated research project. *Prev Vet Med.* 1999;38(2–3):85–100.
8. Ali M, Park JK, von Seidlein L, Acosta CJ, Deen JL, Clemens JD. Organizational aspects and implementation of data systems in large-scale epidemiological studies in less developed countries. *BMC Public Health.* 2006;6:86.
 9. Cummings J, Masten J. Customized dual data entry for computerized data analysis. *Qual Assur.* 1994;3:300–3.
 10. Baumann B. Overcoming obstacles to successful clinical trials through open source. <http://www.appliedclinicaltrials.com/appliedclinicaltrials/article/articleDetail.jsp?id=743711> (2011). Accessed 10 October 2015.
 11. Fegan GW, Lang TA. Could an open-source clinical trial data-management system be what we have all been looking for? *PLOS Med.* 2008;5(3):347–9.
 12. Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, Leizorovicz A, Schade-Brittinger C, Wittenberg M, Voko Z, Gaynor S, Cooney M, Doran P, Maggioni A, Lorimer A, Torres F, McPherson G, Charwill J, Hellström M, Lejeune S. Heterogeneity prevails: the state of clinical trial data management in Europe—results of a survey of ECRIN centres. *Trials.* 2010;11:79.
 13. Lu Z, Su J. Clinical data management: current status, challenges, and future directions from industry perspectives. *Open Access J Clin Trials.* 2010;2:93–105.
 14. Roberts RJ, Musick BS, Olley B, Hall KS, Hendrie HC, Oyediran AB. Data management in a longitudinal cross-cultural study. *Stat Med.* 2000;19(11–12):1645–9.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

