AIDS Care, 2015
Vol. 27, No. S1, 99–107, http://dx.doi.org/10.1080/09540121.2015.1023246

Sankofa pediatric HIV disclosure intervention cyber data management: building capacity in a resource-limited setting and ensuring data quality

Ann Christine Catlin[a], Sumudinie Fernando[a], Ruwan Gamage[a], Lorna Renner[b], Sampson Antwi[c], Jonas Kusah Tettey[b], Kofi Aikins Amisah[c], Tassos Kyriakides[d], Xiangyu Cong[d], Nancy R. Reynolds[e] and Elijah Paintsil[f,g]*

[a]*Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, USA;* [b]*Department of Child Health, Korle-Bu Teaching Hospital, School of Medicine and Dentistry, University of Ghana, Accra, Ghana;* [c]*Department of Child Health, Komfo Anokye Teaching Hospital, School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana;* [d]*Yale Center for Analytical Sciences, Yale School of Public Health, New Haven, CT, USA;* [e]*Yale School of Nursing, West Haven, CT, USA;* [f]*Department of Pediatrics, Yale School of Medicine, New Haven, CT, USA;* [g]*Department of Pharmacology, Yale School of Medicine, New Haven, CT, USA*

(*Received 12 November 2014; accepted 23 February 2015*)

Prevalence of pediatric HIV disclosure is low in resource-limited settings. Innovative, culturally sensitive, and patient-centered disclosure approaches are needed. Conducting such studies in resource-limited settings is not trivial considering the challenges of capturing, cleaning, and storing clinical research data. To overcome some of these challenges, the Sankofa pediatric disclosure intervention adopted an interactive cyber infrastructure for data capture and analysis. The Sankofa Project database system is built on the HUBzero cyber infrastructure (https://hubzero.org), an open source software platform. The hub database components support: (1) data management – the "databases" component creates, configures, and manages database access, backup, repositories, applications, and access control; (2) data collection – the "forms" component is used to build customized web case report forms that incorporate common data elements and include tailored form submit processing to handle error checking, data validation, and data linkage as the data are stored to the database; and (3) data exploration – the "dataviewer" component provides powerful methods for users to view, search, sort, navigate, explore, map, graph, visualize, aggregate, drill-down, compute, and export data from the database. The Sankofa cyber data management tool supports a user-friendly, secure, and systematic collection of all data. We have screened more than 400 child–caregiver dyads and enrolled nearly 300 dyads, with tens of thousands of data elements. The dataviews have successfully supported all data exploration and analysis needs of the Sankofa Project. Moreover, the ability of the sites to query and view data summaries has proven to be an incentive for collecting complete and accurate data. The data system has all the desirable attributes of an electronic data capture tool. It also provides an added advantage of building data management capacity in resource-limited settings due to its innovative data query and summary views and availability of real-time support by the data management team.

Keywords: cyber infrastructure; electronic data capture; dataviews; HIV disclosure; resource-limited setting

Introduction

With the scale up of combination antiretroviral therapy in resource-limited settings, many HIV-infected children are able to survive into adulthood. Yet many of these children do not know they are HIV infected because of caregiver concerns about the child's ability to understand HIV disease, parental sense of guilt and fear of social rejection, and isolation because the child will not keep the diagnosis to him/herself (Wiener, Mellins, Marhefka, & Battles, 2007). Lack of disclosure of HIV status is among the most important and prevalent factors that have been found to negatively affect adherence to treatment in HIV-infected children in sub-Saharan Africa (Kallem, Renner, Ghebremichael, & Paintsil, 2011; Nabukeera-Barungi, Kalyesubula, Kekitiinwa, Byakika-Tusiime, & Musoke, 2007). We have designed an innovative patient-centered intervention approach (Sankofa Project) using a specialist who is familiar with the sociocultural norms of the community and well trained to target modifiable information, motivation, and behavioral skills of caregivers to facilitate their engagement in the process of disclosure in a manner suitable to the needs of the child (Kallem et al., 2011; Vaz et al., 2008; Vreeman et al., 2010). Successful implementation and benefits from the Sankofa Project will depend on acquisition of accurate, reliable, and actionable data.

Building a data support environment for clinical research and trials is a challenging task since the environment must satisfy user and system requirements that encompass all data activities from data collection, update, and audit through data exploration and analysis. The effort is a collaborative one, undertaken jointly by the research, clinical, and data technology groups who

*Corresponding author. Email: elijah.paintsil@yale.edu

must address and resolve many issues together, among them the identification, definition, and relationships of common data elements (CDEs), the design and completeness of case report forms (CRFs), modes of access control for secure data entry and viewing, requirements for review and exploration of data, and specifications for reporting and outcomes analysis. Furthermore, data transmission, access, storage, and security policies must be US Health Insurance Portability and Accountability Act (HIPAA) aligned. Most resource-limited settings lack both the infrastructure and human capacity to manage clinical research and trials effectively (Lang et al., 2010). Thus, there is a need for novel information technology systems and training of research personnel in resource-limited settings to support reliable and quality research (Jaffar et al., 2005; Oliveira & Salgado, 2006). An immediate and cost-effective way of resolving this dilemma is to use a collaborative web-based platform designed specifically for creating and supporting research data infrastructure and capacity in study sites such as those in the Sankofa Project.

The main objective of the data environment for the Sankofa Project was to provide an interactive web-based system for collecting, managing, and exploring clinical data and, in particular, to use features of the system to create innovative audit mechanisms to constantly monitor the consistency and completeness of the data, thus improving quality, accuracy, and reliability of the database. Moreover, the environment provides a model for building in-country data management capacity for other research projects.

Materials and methods

The HUBzero cyber infrastructure

The Sankofa Project database system is built on top of the HUBzero cyber infrastructure (https://hubzero.org). Hub technology was developed at Purdue University through National Science Foundation (NSF) funding, beginning with nanoHUB for the nanotechnology community and then generalizing to the HUBzero platform which offers customizable science gateways for any research community. HUBzero is an open source software platform for creating dynamic, programmable websites that allows users worldwide to deploy research tools, disseminate scientific resources, and distribute educational modules (Hacker et al., 2011; McLennan & Kennell, 2010). Hub technology supports customizable solutions for data collection, exploration, and analytics (Huang et al., 2014; Kuriyan, Catlin, & Reklaitis, 2009; Wang et al., 2013).

Sankofa Project data content, collection, and exploration

A number of tools and instruments for clinical data collection prepared by the Sankofa Project team were mapped to a Protocol Schedule of Events. Data collection tools were designed to collect screening and enrollment information, child demographics, medical history, behavior, medication adherence, and disclosure status. Published and validated instruments are also used in the project to measure social provisions, HIV knowledge, illness perception, HIV stigma, and depression inventory (Beck, Steer, & Brown, 1996; Berger, Ferrans, & Lashley, 2001; Broadbent, Ellis, Thomas, Gamble, & Petrie, 2009; Carey & Schroder, 2002; Cutrona & Russell, 1987). We transformed the form and content of the questionnaires into CRFs using the hub data collection technology and presented the forms for longitudinal data entry according to a customized knowledge-based events scheduler.

We also designed "dataviews" needed to browse, search, explore, audit, drill-down, analyze, and export the collected data. The dataview capability is the most sophisticated, flexible, and powerful component of the hub data technology, allowing us to create customized "windows" into the database in any way that serves the needs of the Sankofa Project team. Clinicians need to browse and review raw patient data; statisticians need advanced analytics, with raw, derived, aggregate, and integrated data. Dataviews were created for various viewpoints, with data at the required of detail.

Security and privacy of "Limited Data Set" electronic protected health information (ePHI)

The information collected from Sankofa participant is classified as a HIPAA "Limited Data Set" which requires that the hub infrastructure, data technology, and medical database be HIPAA aligned. The cceHUB gateway for the Sankofa Project (https://ccehub.org) provides all required safeguards for user authentication, access establishment, monitoring, audit, integrity, backup, and disaster recovery. Every Sankofa Project team member is a registered user of cceHUB, with authorization granted by the project's management team to enter/contribute data, explore data, or both. Data entry/contribution is restricted to individuals at the two participating study sites in Ghana; data contributors are authenticated by site for all access to the database, both for data collection and dataviewing. Data privileges are tightly controlled, with group-based restriction covering all modes of database access. Team members belong to one or more "invite only" hub groups, which were established for the assignment of data privileges and for sharing restricted dataviews for exploring the database.

Patient information collected for this project is de-identified by auto-assigning alphanumeric identifiers for newly entered screened and enrolled patients. Data contributors are responsible for separately maintaining linkage between participants and Sankofa identifiers. All ePHI are encrypted for storage and decrypted on the fly for shared access and viewing.

The study protocol was approved by the Institutional Review Boards of Yale School of Medicine, Purdue University, University of Ghana School of Medicine and Dentistry, and Komfo Anokye Teaching Hospital. All participants gave their written informed consent before participation in the study.

Results

Sankofa project database development

The Sankofa database was designed and developed from October 2012 to January 2013 and opened for participant screening and enrollment on 1 February 2013 at our web-based portal https://ccehub.org/sankofa.

The design and development of customized hub databases is a well-organized process which is based on the "building blocks" of our technology: (1) data management – the "databases" component creates, configures, and manages database access, backup, repositories, applications, and access control; (2) data collection – the "forms" component supports the building of customized web CRFs that incorporate CDEs and include tailored form submit processing to handle error checking, data validation, and data linkage as the data are stored to the database; and (3) data exploration – the "dataviewer" component provides powerful methods for users to view, search, sort, navigate, explore, map, graph, visualize, aggregate, drill-down, compute, and export data from the database (Figure 1).

The data collection component is exercised heavily in the first phase of database building, and this effort is tightly connected to the design of the database schema. The database developers worked with the researchers to acquire a detailed and comprehensive understanding of (1) all elements of the research data, (2) all tools, instruments, and processes used in data gathering, and (3) all relevant information about patients, drugs, adherence, and clinic policies. This resulted in a system specification for 25 separate web forms containing more than 300 data elements, some with complex relationships between elements from different forms, some with variable dependencies on the drug knowledge base (which we built into the database and made available to the forms), and some with specialized processing at submission time to maintain data consistency, in particular with respect to the primary outcome (HIV disclosure). We had to ensure flexibility of design, as some of the questionnaires and submit processing requirements changed over the course of the project. All forms maintained attributes for participant ID codes (assigned at screening, enrollment) and data entry tagging (assigned for site, contributor, and submission time stamp).

In addition to a customized "form definition" for each web form, the data collection process required a customized, interactive, and graphical scheduler to track the status of forms submission for each patient across the timeline of study weeks (most forms are longitudinal). To ease form entry for the clinical team, our scheduler

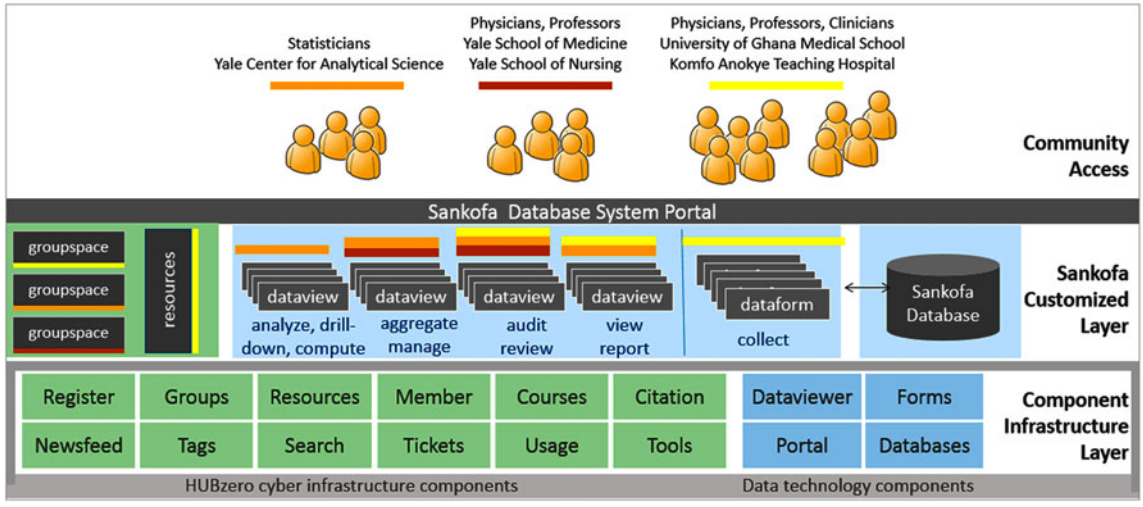


Figure 1. Sankofa database management structure. The infrastructure layer shows the components of HUBzero and the data technology that support the Sankofa Project. In the customized layer are the forms, views, groups, and database created using the components. This diagram illustrates the global nature of the interactions between clinical and research group, with the role-based access and sharing of de-identified patient data through a web-based portal to a centralized research database established on the collaborative HUBzero cyber infrastructure.

interface detects the site and corresponding participant list of the submitter, then guides the data entry workflow for the selected participant – strictly enforcing data entry for forms required at entry week, off-study week, and the critically important disclosure period, but providing flexibility for forms and study week selection during the standard weeks period (since these requirements can shift and vary depending on the occurrence of disclosure).

In the second phase of the project, the data exploration component was the main focus of the developers, since it is used to support ongoing efforts to understand and analyze data stored in the Sankofa database. An initial collection of dataviews was generated to aid clinical teams and biostatisticians in navigating and reviewing the raw data. Our views are tabular Excel-like presentations of the patient data and are generated instantly to display data stored for each of the forms (one default dataview per form). These dataviews provide comprehensive and feature-rich browsing of patient data through search, sort, "Google-like" text filtering on columns, numeric range checking, and clickable drill-down on patient IDs to move across forms.

The basic views are valuable for data verification and checking and are in constant use (Figure 2). However, the real power of the data exploration component is its support for easily customizable dataviews that can (1) join data across the entire database for specific types of investigations, (2) compute and aggregate on data for auditing or analytics, and (3) offer drill-down at any level to link targeted research data together. The data exploration component provides a simple, high-level language (the "data definition") to easily create complex, sophisticated views that can be used to answer complex questions about the data.

Dataviews for data quality, accuracy, and completeness

A major goal of our project was to use dataviews to provide reliable, real-time, interactive interfaces that support continuous monitoring of the research data to ensure its accuracy and completeness. The measurement and reporting on Sankofa data completeness consists of three principal tasks: (1) audit the data from each submitted form for accuracy and completeness, including the tracking and aggregation of data elements marked by clinicians as "missing" or "patient refused to answer" (pra); (2) verify the consistency of the data stored to the database, in particular when different sources are describing the same data element (e.g., caregiver vs. health-care provider vs. clinician); and (3) monitor the progress of forms submitted for each patient to ensure that the collection of forms is on track and complete, according to the requirements of the events schedule timeline.

Auditing data

To quantify the extent of "missing" and "pra" used as answers across the submitted forms, computational dataviews were created to aggregate the counts of "missing" and "pra" responses for each data element in all 25 forms for the entire patient population (Figure 3). These views were continuously available, computing aggregations in real time, and offering the usual powerful and user-friendly exploration features.

Verifying data consistency

The consistency and correctness of the database data is assessed by a validation process that is invoked when forms are submitted. However, some key driving factors – like date of disclosure – are described by different sources and often differ. Since the disclosure date is used to

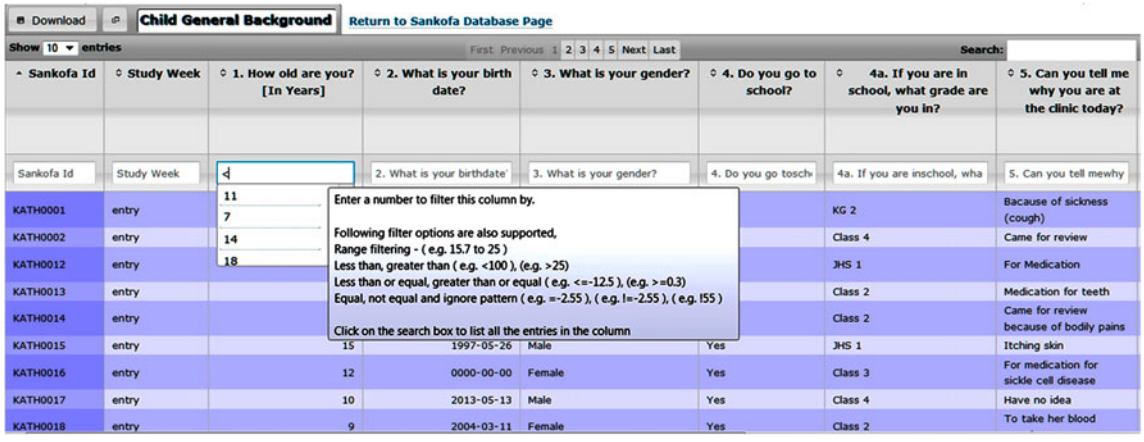


Figure 2. Dataview of baseline data at enrollment. Basic dataviews present raw data from a form. Data are controlled so that sites in Ghana can see only their own patients, statisticians can see all patients, and researchers from Yale cannot access the view. Columns can be sorted and dataviews are searchable on all columns using numeric or text-based features. Exported data are based on the filtered display.

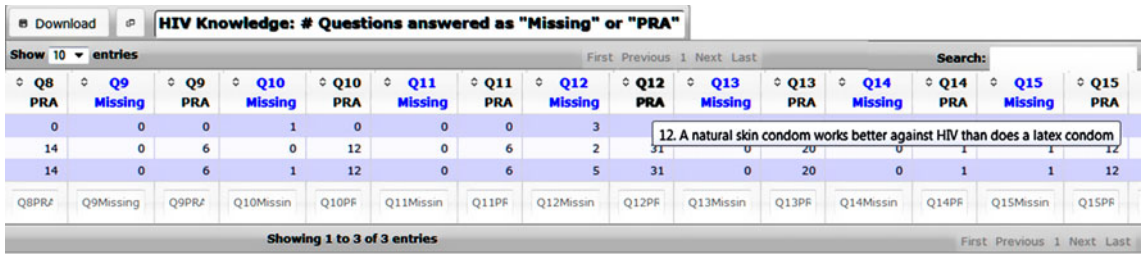


Figure 3. Audit dataview for missing and pra questions. Aggregate data for specific answers to questions are shown in this dataview. Note that Missing and PRA have specific meanings (different from "blank" fields), and this view is used to understand how patients are responding to the questions, so that decisions can be made about the effectiveness of the question formulation and questionnaire format.

determine how participant care should proceed for subsequent study weeks (with events data captured into the database), the clinician is tasked with responsibility for precise identification of the disclosure week to the database. A special dataview was created to ensure the correctness of disclosure status. This view made the disclosure status determination straightforward across all patients, by identifying current status and related disclosure information according to all sources for easy comparison and review (Figure 4).

Monitoring data completeness

Progress in submitting patient forms across the study weeks timeline is a critical measure of data completeness, and the tracking and assessment of form submission is one of the key tasks assigned to the Sankofa management group. To make the completeness assessment easy, interactive, and real-time, we created a graphical display of aggregate submitted form totals on the study week timeline for clinicians, researchers, and biostatisticians (Figure 5). The display identifies the total number of forms submitted at each study week, with clickable drill-down to display which participants have forms submitted. The aggregate numbers can be used for reviewing process and policy to ensure the study is progressing as expected.

Dataviews for statistical and analytical data exploration

The sophisticated computational features of the data exploration component can also be demonstrated by dataviews created to support the Sankofa biostatistics group. We created many useful views for Sankofa biostatisticians, including (1) a way to investigate the data to ensure its quality and readiness for statistical analysis and (2) a way to download precisely the content, formats, and calculations that were needed as input for their statistical analysis tools (e.g., SAS). We will describe two requests and the corresponding views.

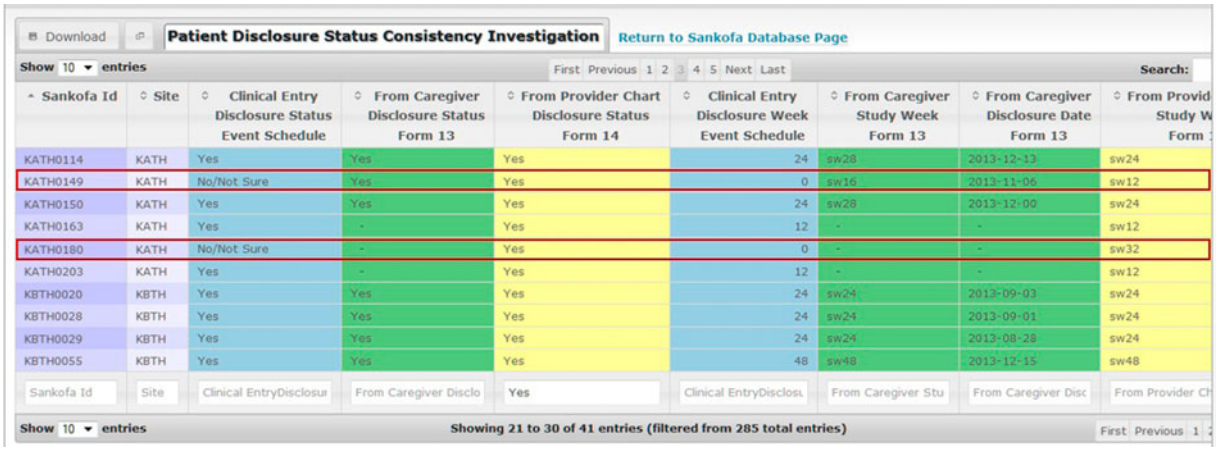
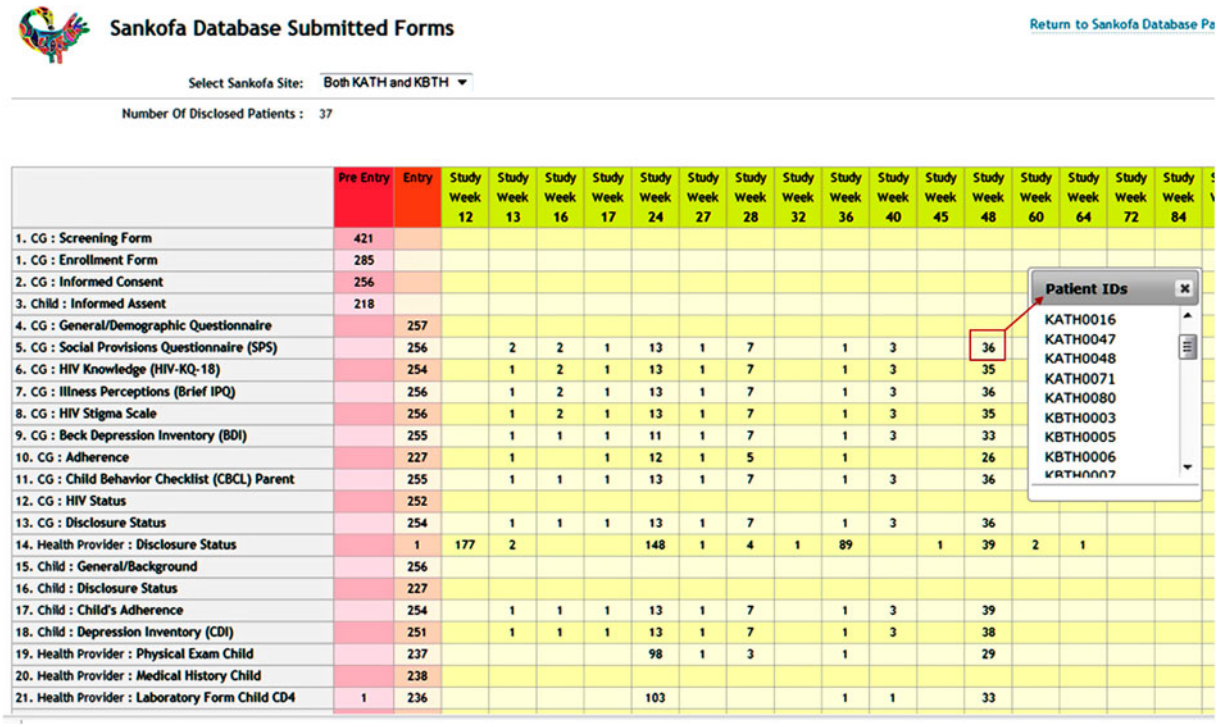


Figure 4. Consistency of HIV disclosure status. Audit dataviews were created to ensure consistency and accuracy of the primary outcome data. In this view, the HIV disclosure date identified by the clinician can be checked against values provides by different sources on different forms, with status, source, and date color-coded for easy checking. This view indicates that the clinician (blue) should update the site data according to health-care provided information (yellow). Caregiver (green) and Provider data are allowed to differ.



	Pre Entry	Entry	Study Week 12	Study Week 13	Study Week 16	Study Week 17	Study Week 24	Study Week 27	Study Week 28	Study Week 32	Study Week 36	Study Week 40	Study Week 45	Study Week 48	Study Week 60	Study Week 64	Study Week 72	Study Week 84
1. CG : Screening Form	421																	
1. CG : Enrollment Form	285																	
2. CG : Informed Consent	256																	
3. Child : Informed Assent	218																	
4. CG : General/Demographic Questionnaire		257																
5. CG : Social Provisions Questionnaire (SPS)		256		2	2	1	13	1	7		1	3		36				
6. CG : HIV Knowledge (HIV-KQ-18)		254		1	2	1	13	1	7		1	3		35				
7. CG : Illness Perceptions (Brief IPQ)		256		1	2	1	13	1	7		1	3		36				
8. CG : HIV Stigma Scale		256		1	2	1	13	1	7		1	3		35				
9. CG : Beck Depression Inventory (BDI)		255		1	1	1	11	1	7		1	3		33				
10. CG : Adherence		227		1		1	12	1	5		1			26				
11. CG : Child Behavior Checklist (CBCL) Parent		255		1	1	1	13	1	7		1	3		36				
12. CG : HIV Status		252																
13. CG : Disclosure Status		254		1	1	1	13	1	7		1	3		36				
14. Health Provider : Disclosure Status		1	177	2			148	1	4	1	89		1	39	2	1		
15. Child : General/Background		256																
16. Child : Disclosure Status		227																
17. Child : Child's Adherence		254		1	1	1	13	1	7		1	3		39				
18. Child : Depression Inventory (CDI)		251		1	1	1	13	1	7		1	3		38				
19. Health Provider : Physical Exam Child		237					98	1	3		1			29				
20. Health Provider : Medical History Child		238																
21. Health Provider : Laboratory Form Child CD4	1	236					103				1	1		33				

Figure 5. The events scheduler. Tracks the flow of submitted forms across the study week timeline. It guides clinical teams in data entry workflow, aggregates form totals by patient for data completeness, and provides drill-down to audit which forms and patients are missing data. This supports the collection of forms on a longitudinal basis according to patient and site requirements – the forms correspondence with events is *not* fixed a priori. In this audit view, tracking for forms and study weeks show aggregated form submit totals across the study, with click-on drill-down to patient IDs. A site-specific tracker checks by selected patient, with drill-down to the forms to edit.

Monitoring the progress of the study

The biostatisticians were required to submit template reports on the status of data flowing into the database. In most projects, these reports are generated infrequently due to the effort required to gather and compute data in the appropriate formats. For Sankofa, we created dataviews to present the data flow status for real-time, always available database monitoring. Summary dataviews report on important statistics (e.g., means, frequencies) for characteristics and categories of participant and caregiver populations, presenting data that merge and operate on both raw and derived data.

Automating statistical report generation for raw data

Our basic dataviews for raw data are forms-based, with additional customization to support linkage between forms as a patient drill-down. These 25 dataviews are principally used for reviewing, searching, and exploring patient data, and each dataview can be downloaded into a spreadsheet containing the raw patient data from particular form. The biostatisticians needed a dataview (and its built-in spreadsheet download) that combined data into much larger views that encompassed raw data from many forms. The dataviews would then be downloaded weekly, SAS macros would be run to generate the reports, and the spreadsheets and reports would be archived on a weekly basis. We created three "Large Dataviews" for raw data in "wide format," combining the data from the forms as needed to satisfy the needs of automated reporting. Some additional dataviews were also created to present raw data in different formats to satisfy requirements for SAS input.

To clarify the power of the dataviews: the database schema and content depend only on the data elements and relationships required for the forms, there is no schema support for computations, aggregations, or analysis of any kind. It is the dataview itself that contains the needed joins, computations, and analytics to produce *any requested web-based view of combined raw and derived data*. These are created easily and immediately using the "data definition" capability of the data exploration component, with powerful search, sort, filter, drill-down, and export features built into every view. Dataview analytics can be used to transform a database system into a searchable knowledge base.

Discussion

Ironically, resource-limited settings have the most disease burden and the greatest potential to benefit from effective intervention from well-designed clinical research. Fewer studies are undertaken in resource-limited settings due to the daunting challenges of capturing, cleaning, extracting, and storing clinical research data (Fegan & Lang, 2008; Lang et al., 2010; Naresh et al., 2011). To overcome some of these challenges, Sankofa adopted an interactive cyber infrastructure for data capture and analysis. Moreover, the HUBzero cyber infrastructure offers a collaborative solution by giving in-country research personnel interactive access to: (1) training in good clinical practices, (2) guidelines and tools for data acquisition, and (3) built-in analytics for the acquired data. It also offers collaborative group spaces where users can post member-restricted resources, create wiki pages to add useful content, ask questions, and get answers. To the best of our knowledge, this is the first time an electronic data capture (EDC) tool with the capabilities of the HUBzero cyber infrastructure has been deployed in sub-Saharan Africa for the data management of a clinical research.

An ideal EDC tool should have the following key attributes (Franklin, Guidry, & Brinkley, 2011; Harris et al., 2009): (1) collaborative access to data, (2) user authentication and role-based security, (3) user-friendly CRFs, (4) real-time data validation and integrity checks, (5) data attribution and audit capabilities, (6) protocol document storage and sharing, (7) central data storage and backups, (8) data export function for common statistical packages, and (9) data import function. Several EDC tools have been developed and deployed in the last decade to meet these challenges ("Archive of Visitrial website, 2006"; "Caisis," 2011; Harris et al., 2009; "LabKey," 2011; "OpenClinica from Akaza Research," 2011; "TrialDB website," 2007; "University of Washington Catalyst Web Tools," 2011). In a head-to-head comparison of the functionality of available EDC tools including Research Electronic Data Capture (REDCap), Catalyst Web Tools, OpenClinica, Caisis, and LabKey, Franklin et al. found REDCap to have a very clear advantage over most of the EDC tools due to its extensive tutorials and online training material (Franklin et al., 2011). Though OpenClinica has more functionality, for example, in complex CRF design and site management, there is less documentation, and what is available is written in technical language. They concluded that the limitations of Catalyst Web Tools were too great to manage a clinical trial. They also found that Caisis and LabKey did not meet specific user needs in a clinical trial.

REDCap was developed and deployed at Vanderbilt University and is available to a wide consortium of domestic and international research collaborations. We choose to compare our system to REDCap since it is the most widely used EDC tool. First, while REDCap offers a user-friendly system for building data entry forms, only built-in features of the official base code can be used. The base code does not allow end users to customize feature behavior, form operation, or form submit processing, though requests to the REDCap informatics core can be made for future consideration. Our data collection system is flexible, with customization and changes to feature behavior and processing allowed at any time during development or production mode. For example, data entry fields can display selections based on existing content of the database and user access privileges. Data entered by users to a form can alter menu options available for subsequent use of the form (such as dynamically adding new types of adverse events). Furthermore, forms submit processing can operate on any data element in the database, not just those submitted from the form. Second, REDCap data are stored in a single database table with no relational structure. This can be inefficient from a data storage stand point. Our database has a relational structure, making data storage and exploration more flexible. Third, REDCap does not prevent submission of data that is out of range. It generates warning messages, but these can be overridden. Our submission validation is more robust and can be customized as needed to satisfy data quality requirements. Fourth, although the data collection function in REDCap is popular among users, our data exploration capability is significantly more advanced. Our dataviews can display raw data from across the database merged with derived data, then grouped and filtered as needed to present unique entries for investigation and comparison. Different dataviews can be linked together using "filter," "drill-down" and "more info" features. Every dataview supports powerful interactive data exploration features, such as numeric value and range searches or Google-like text searches on multiple columns or across the dataview. This means that all searches on a particular collection of data occur on a single dataview, unlike REDCap where a new report must be built for every new query. We can create statistical dataviews applying any mathematical function to any combination of data elements, and we can even launch external tools from the dataview with data passed in from the database. Dataviews can instantly display maps with markers listing metadata for any data associated with location coordinates, and we can generate graphs for any user-selected (x, y) data. Thus, our system provides not only data collection but also feature-rich data exploration, all established on a single collaborative cyber platform.

One of REDCap's principal features is that research groups themselves create and control their own forms and reports, with no need for IT developers – this is a valuable and significant achievement. However, a

HIPAA-aligned platform for installation and use of REDCap must still be established, with local IT for support. Research groups must still defer to REDCap support teams for changes to the official base code, forego customized submit processing, and depend on future REDCap development for desired new features that are needed for more complex data collection. The global REDCap consortium does offer plug-ins with extended capabilities, but these must be individually installed and are not guaranteed to work as needed. Most significantly, the REDCap base code system is primarily a vehicle for data collection, whereas hub databases provide a vehicle for data exploration, analysis, and knowledge building.

With easy to use data elements, CRFs, and clarified data collection sequences, the Sankofa cyber data management tool has fully supported a user-friendly, secure, and systematic collection of all data envisioned for carrying out the research objectives. We currently have screened more than 400 patients and enrolled nearly 300 patients, with tens of thousands of data elements collected to describe demographic, medical, and psychosocial events during the course of the study period. The dataviews have successfully supported all data exploration and analysis needs of the Sankofa Project. Moreover, the ability of the sites to query and view data summaries has proven to be an incentive for collecting complete and accurate data. The data management team is available in real time to help with troubleshooting and providing solutions to the sites. In so doing, the data entry team and project coordinators continue to enrich their capacity for data management.

Cost will continue to be a barrier to all aspects of research. With the deployment of EDC tools for clinical trials, substantial reduction in cost has been demonstrated. The cost savings have accrued from less time required for data entry; less redundancy in the data recording; built-in error checking that results in fewer enrollment of invalid patients; easier access to database; the ability to clean, analyze, and report more accurate data; and reduction in clinical trial duration and time to database closure (Bart, 2003; Welker, 2007). The cost of our product has been within the budget module of a National Institute of Health R01 funding mechanism. The only limitation will be the deployment of the product in resource-limited settings where there are frequent power outages and intermittent Internet access. We circumvented this by providing mobile Internet modems, which is available in most resource-limited settings, for data collection during periods when broadband Internet is unavailable.

In conclusion, the Sankofa Project data system has all the desirable attributes of an EDC tool and more. It will ensure an accurate and reliable database. Moreover, it is cost-effective and can be deployed for clinical trials and translation research activities in resource-limited settings. It also provides an added advantage of building data management capacity in resource-limited settings due to its innovative data query and summary views and availability of real-time support by the data management team.

Acknowledgments

We thank the Sankofa Project caregiver and child dyads for their participations. We are grateful to the staff at Pediatric AIDS Clinics at Korle-Bu and Komfo Anokye Teaching Hospitals and the Ghana–Yale Partnership for Global Health for their support.

Members of the Sankofa Project Team. Elijah Paintsil, Nancy Reynolds, Tassos Kyriakides, Xiangyu Cong, and Yram Foli, Yale University, USA; Lorna Renner, Margaret Lartey, Angela Ofori-Atta, Jonas Kusah Tetteh, Joyceline Assimeng, Obedia Akweley Seaneke, Dramani Yakubu, and Kevin Bonsu, Korle-Bu Teaching Hospital, Accra, Ghana; Sampson Antwi, Kofi Aikins Amissah, Anthony Enimil, Amina Alhassan, and Irene Pokuaa Ofori, Komfo Anokye Teaching Hospital, Kumasi, Ghana; Ann Christine Catlin, Sumudinie Fernando, Ruwan Egoda Gamage, Ruchith Fernando, and Sudheera Fernando, Purdue University, USA.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was made possible by grant from NIH/NICHD [grant number R01HD074253]. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official view of NIH.

References

Archive of Visitrial website. (2006) Retrieved February 5, 2014, from http://web.archive.org/web/20070329221856/http://devctr.visitrial.com

Bart, T. (2003). Comparison of electronic data capture with paper data collection – Is there really an advantage? In *Business briefing: Pharmatech* (pp. 1–4) [World Markets Series].

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory: Texas psychological corporation*. New York, NY: Harcourt Brace & Co.

Berger, B. E., Ferrans, C.E., & Lashley, F. R. (2001). Measuring stigma in people with HIV: Psychometric assessment of the HIV stigma scale. *Research in Nursing and Health*, *24*, 518–529. doi:10.1002/nur.10011

Broadbent, E., Ellis, C. J., Thomas, J., Gamble, G., & Petrie, K. J. (2009). Can an illness perception intervention reduce illness anxiety in spouses of myocardial infarction patients? A randomized controlled trial. *Journal of Psychosomatic Research*, *67*(1), 11–15. doi:10.1016/j.jpsychores.2008.11.006

Caisis. (2011). Retrieved February 5, 2014, from http://www.caisis.org

Carey, M. P., & Schroder, K. E. E. (2002). Development and psychometric evaluation of the brief HIV knowledge questionnaire. *AIDS Education and Prevention*, *14*, 172–182. doi:10.1521/aeap.14.2.172.23902

Cutrona, C. E., & Russell, D. W. (1987). The provisions of social relationships and adaptation to stress. *Advances in Personal Relationships*, *1*, 37–67.

Fegan, G. W., & Lang, T. A. (2008). Could an open-source clinical trial data-management system be what we have all been looking for? *PLoS Medicine*, *5*(3), e6. doi:1740-7753(2006)003[0385:DAOADC]2.0.CO;2

Franklin, J. D., Guidry, A., & Brinkley, J. F. (2011). A partnership approach for Electronic Data Capture in small-scale clinical trials. *Journal of Biomedical Informatics*, *44*(Suppl. 1), S103–S108. doi:10.1016/j.jbi.2011.05.008

Hacker, T. J., Eigenmann, R., Bagchi, S., Irfanoglu, A., Pujol, S., Catlin, A., & Rathje, E. (2011). The NEEShub cyberinfrastructure for earthquake engineering. *Computing in Science and Engineering*, *13*(4), 67–78. doi:10.1109/MCSE.2011.70

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*, 377–381. doi:10.1016/j.jbi.2008.08.010

Huang, J., Ahmad, U., Antonicelli, A., Catlin, A. C., Fang, W., Gomez, D., … Detterbeck, F. (2014). Development of the international thymic malignancy interest group international database: An unprecedented resource for the study of a rare tumor. *Journal of Thoracic Oncology*, *9*(10), 1573–1578. doi:10.1097/JTO.0000000000000269

Jaffar, S., Govender, T., Garrib, A., Welz, T., Grosskurth, H., Smith, P. G., … Bennish, M. L. (2005). Antiretroviral treatment in resource-poor settings: Public health research priorities. *Tropical Medicine and International Health*, *10*, 295–299. doi:10.1111/j.1365-3156.2005.01390.x

Kallem, S., Renner, L., Ghebremichael, M., & Paintsil, E. (2011). Prevalence and pattern of disclosure of HIV status in HIV-infected children in Ghana. *AIDS and Behavior*, *15*, 1121–1127. doi:10.1007/s10461-010-9741-9

Kuriyan, K., Catlin, A. C., & Reklaitis, G. V. (2009). pharmaHUB: Building a virtual organization for pharmaceutical engineering and science. *Journal of Pharmaceutical Innovation*, *4*(2), 81–89. doi:10.1007/s12247-009-9061-7

LabKey. (2011). Retrieved February 5, 2014, from http://www.labkey.com

Lang, T. A., White, N. J., Hien, T. T., Farrar, J. J., Day, N. P. J., Fitzpatrick, R., … Marsh, K. (2010). Clinical research in resource-limited settings: Enhancing research capacity and working together to make trials less complicated. *PLoS Neglected Tropical Diseases*, *4*(6), e619. doi:10.1371/journal.pntd.0000619.g001

McLennan, M., & Kennell, R. (2010). HUBzero: A platform for dissemination and collaboration in computational science and engineering. *Computing in Science and Engineering*, *12*(2), 48–53. doi:10.1109/MCSE.2010.41

Nabukeera-Barungi, N., Kalyesubula, I., Kekitiinwa, A., Byakika-Tusiime, J., & Musoke, P. (2007). Adherence to antiretroviral therapy in children attending Mulago Hospital, Kampala. *Annals of Tropical Paediatrics*, *27*(2), 123–131. doi:10.1179/146532807X192499

Naresh, K. N., Raphael, M., Ayers, L., Hurwitz, N., Calbi, V., Rogena, E., … Leoncini, L. (2011). Lymphomas in sub-Saharan Africa – What can we learn and how can we help in improving diagnosis, managing patients and fostering translational research? *British Journal of Haematology*, *154*, 696–703. doi:10.1111/j.1365-2141.2011.08772.x

Oliveira, A. G., & Salgado, N. C. (2006). Design aspects of a distributed clinical trials information system. *Clinical Trials*, *3*, 385–396. doi:10.1177/1740774506069156

OpenClinica from Akaza Research. (2011). Retrieved February 5, 2014, from http://www.openclinica.com

TrialDB website. (2007). Retrieved February 5, 2014, from http://ycmi.med.yale.edu/trialdb/

University of Washington Catalyst Web Tools. (2011). Retrieved February 5, 2014, from http://www.washington.edu/1st/web_tools

Vaz, L., Corneli, A., Dulyx, J., Rennie, S., Omba, S., Kitetele, F., … Behets, F. (2008). The process of HIV status disclosure to HIV-positive youth in Kinshasa, Democratic Republic of the Congo. *AIDS Care*, *20*, 842–852. doi:10.1080/09540120701742276

Vreeman, R. C., Nyandiko, W. M., Ayaya, S. O., Walumbe, E. G., Marrero, D. G., & Inui, T. S. (2010) The perceived impact of disclosure of pediatric HIV status on pediatric antiretroviral therapy adherence, child well-being, and social relationships in a resource-limited setting. *AIDS Patient Care and STDs*, *24*, 639–649. doi:10.1089/apc.2010.0079

Wang, T., Alston, K., Wassgren, C., Mockus, L., Catlin, A. C., Fernando, S., & Fernando, S. R. (2013). The creation of an excipient properties database to support quality by design (QbD) formulation development. *American Pharmaceutical Review*, *16*(4), 16–25.

Welker, J. A. (2007). Implementation of electronic data capture systems: Barriers and solutions. *Contemporary Clinical Trials*, *28*, 329–336. doi:10.1016/j.cct.2007.01.001

Wiener, L., Mellins, C. A., Marhefka, S., & Battles, H. B. (2007). Disclosure of an HIV diagnosis to children: History, current research, and future directions. *Journal of Developmental and Behavioral Pediatrics*, *28*, 155–166. doi:10.1097/01.DBP.0000267570.87564.cd