

RESEARCH

Open Access



The Ugandan sickle Pan-African research consortium registry: design, development, and lessons

Mike Nsubuga^{1,2,3,4}, Henry Mutegeki^{3,4,5}, Daudi Jjingo^{3,4,5}, Deogratias Munube⁶, Ruth Namazzi⁶, Robert Opoka^{6,7,9}, Philip Kasirye⁶, Grace Ndeezi⁶, Heather Hume⁸, Ezekiel Mupere⁶, Grace Kebirungi^{3,4}, Isaac Birungi⁹, Jack Morrice¹⁰, Mario Jonas¹⁰, Victoria Nembaware¹⁰, Ambroise Wonkam¹¹, Julie Makani¹² and Sarah Kiguli^{6*}

Abstract

Background Sub-Saharan Africa bears the highest burden of sickle cell disease (SCD) globally with Nigeria, Democratic Republic of Congo, Tanzania, Uganda being the most affected countries. Uganda reports approximately 20,000 SCD births annually, constituting 6.67% of reported global SCD births. Despite this, there is a paucity of comprehensive data on SCD from the African continent. SCD registries offer a promising avenue for conducting prospective studies, elucidating disease severity patterns, and evaluating the intricate interplay of social, environmental, and genetic factors. This paper describes the establishment of the Sickle Pan Africa Research Consortium (SPARCo) Uganda registry, encompassing its design, development, data collection, and key insights learned, aligning with collaborative efforts in Nigeria, Tanzania, and Ghana SPARCo registries.

Methods The registry was created using pre-existing case report forms harmonized from the SPARCo data dictionary and ontology to fit Uganda clinical needs. The case report forms were developed with SCD data elements of interest including demographics, consent, baseline, clinical, laboratory and others. That data was then parsed into a customized REDCap database, configured to suit the optimized ontologies and support retrieval aggregations and analyses. Patients were enrolled from one national referral and three regional referral hospitals in Uganda.

Results A nationwide electronic patient-consented registry for SCD was established from four regional hospitals. A total of 5,655 patients were enrolled from Mulago National Referral Hospital (58%), Jinja Regional Referral (14.4%), Mbale Regional Referral (16.9%), and Lira Regional Referral (10.7%) hospitals between June 2022 and October 2023.

Conclusion Uganda has been able to develop a SCD registry consistent with data from Tanzania, Nigeria and Ghana. Our findings demonstrate that it's feasible to develop longitudinal SCD registries in sub-Saharan Africa. These registries will be crucial for facilitating a range of studies, including the analysis of SCD clinical phenotypes and patient outcomes, newborn screening, and evaluation of hydroxyurea use, among others. This initiative underscores the potential for developing comprehensive disease registries in resource-limited settings, fostering collaborative, data-driven research efforts aimed at addressing the multifaceted challenges of SCD in Africa.

*Correspondence:
Sarah Kiguli
skwalube@yahoo.co.uk

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Sickle cell disease, SPARCo, Disease registry

Introduction

Sickle cell disease (SCD) is a public health challenge with a significant burden globally, particularly affecting sub-Saharan Africa, where 75% of the disease occurs [1]. That comparatively higher prevalence is largely driven by malaria, against which persons with the sickle cell trait have higher resistance [2]. Indeed, the mutation causing SCD arose in Africa thousands of years ago and was propagated within human populations to help protect against malaria, a historically major cause of death on the continent [3]. Characterized by a group of inherited red blood cell disorders, sickle cell anemia, the most severe form of SCD, results from the homozygous inheritance of the sickle hemoglobin gene. This genetic alteration leads to the production of abnormally shaped red blood cells, precipitating a cascade of phenotypical clinical complications ranging from acute pain, chronic pain and severe infections to stroke and organ damage [4].

Each year, more than 400,000 babies are born with SCD globally, with the majority, approximately 300,000, affected by sickle cell anemia [1, 5]. These numbers are expected to double by 2050 [5], underscoring the urgent need for improved care, research, and prevention strategies. In Uganda, SCD is a public health priority, with approximately 20,000 babies born with sickle cell anemia annually, constituting roughly 6.67% of reported global SCD births [6].

Although sub-Saharan Africa bears the highest burden of SCD worldwide, there is a lack of comprehensive data from well-designed longitudinal studies on SCD essential for understanding the disease's epidemiology, progression, and the interplay of genetic and environmental factors influencing its cause [7]. Recently, there has been an increase in SCD research from Africa [8–10], signaling a growing recognition of the disease's burden and an increased focus on addressing it. However, most studies are single-center, employ inconsistent data collection methods, and are not longitudinal. In order to better understand the epidemiology, disease progression, and disease modifiers of SCD, there is a need to conduct well-designed multi-center, long-term longitudinal cohort studies in Africa, backed by appropriately designed registries.

SCD registries are particularly useful for long-term follow-up studies that aim to understand the patterns of disease severity, and the social, environmental determinants of outcomes in patients with SCD in Africa. The establishment of centralized, national, electronic, SCD registries consented by patients in Africa holds the potential to revolutionize SCD research and treatment. Such a registry not only promises a deeper understanding of the

disease but also serves as the cornerstone for the development of personalized healthcare strategies tailored to the unique and diverse needs of SCD patients [1].

Uganda joined the Sickle Pan Africa Research Consortium (SPARCo) in 2020 [11]. SPARCo is a collaborative initiative involving 6 countries (Ghana, Nigeria, Tanzania, Mali, Zambia, Zimbabwe, and Uganda) whose aim is to improve outcomes of persons with SCD through multi-center research and regionally adapted evidence based clinical care in Africa. Here, we present an in-depth description of the development of an electronic patient consented registry in Uganda. We also present an analysis of the patient enrollment and management thereby contributing valuable insights and laying the foundation for future research and intervention strategies tailored to combat SCD's profound impact in sub-Saharan Africa.

Methods

Study sites

The registry was developed at Mulago National Referral Hospital Sickle cell clinic in Kampala, the capital city of Uganda, which is one of the most populous areas with 1.6 million inhabitants [12]. The Mulago Sickle cell clinic is the largest and oldest clinic in Uganda and the East African region, and serves mainly the central region, but receives patients from the rest of the country. Subsequently, the registry was expanded to three additional Regional Referral Hospitals: Jinja Regional Referral Hospital in Eastern Uganda, Mbale Regional Referral Hospital and Lira Regional Referral Hospital. The geographic distribution of the program's catchment areas thus extends across three of four major regions in the country, all of which possess a high prevalence of SCD [6]. The Ministry of Health (MOH) of Uganda was involved as key stakeholders from the inception of the registry, with regular updates and participation in the Advisory Committee to ensure sustainability. Furthermore, the health workers who were responsible for enrolling patients were employed by the MOH, which helped promote the registry's long-term viability.

Study design

The registry utilized a longitudinal study design, wherein data was systematically collected from each patient during every follow-up visit.

Database design

The SPARCo Uganda registry was developed using a comprehensive approach, beginning with the adoption of standardized ontologies for SCD under the coordination of the Sickle Africa Data Coordinating Center (SADaCC)

[13, 14]. Adjustments and customizations of these were then applied to suit local context, both by way of collected data types and local clinical standards (Fig. 1). The harmonization and mapping were done in Python and scripts have been added in the code section. These optimized ontologies facilitated structured data representation and standardized data capture. Following this, a Case Report Form (CRF) was developed, aligning data items from various sources, including existing forms at Mulago National Referral Hospital, to ensure consistency and accuracy in data collection. This process was then tailored to fit specific clinical processes at each site. We integrated Research Electronic Data Capture (REDCap), a versatile web-based platform optimized for modern scientific data collection [15, 16]. Scripts were developed utilizing its built-in validation rules and automated data quality checks including blank value checks, incorrect values for calculated fields checks and field validation to ensure the integrity of research data.

Implementation of registry deployment and mobile app utilization for data collection

The deployment of the registry was set up on a Linux server running CentOS 8 and housed at the African Center for Excellence in Bioinformatics and Data Intensive Sciences (ACE) [17]. Single Sign-On authentication was integrated to enhance security and multi-site access. This approach simplified user access across different sites. Additionally, the REDCap mobile application was implemented, enabling data entry personnel at various hospital sites to efficiently collect data. This was also configured to allow for data collection both in online and offline modes, ensuring uninterrupted data capture (Fig. 2).

Participant consent procedures

The study procedures were explained to both patients and/or their caregivers. Prior to enrollment in the registry, individuals who consented to participate were required to sign a written informed consent form, indicating their approval. In instances where patients were below 18 years of age, informed consent was obtained from their caregivers. For patients aged 18 years and older, consent was directly obtained from the patients themselves. Patients aged between 8 and 17 years were provided with assent forms, in addition to written informed consent from their caregivers.

Data entrant training

Training workshops were conducted at each site, where data managers, coordinators, and clerks were trained on completing CRFs and inputting data electronically. These sessions were designed to ensure uniformity and high-quality data collection across all sites.

Patient enrollment and data collection

Using CRF, demographic data, including date of birth/age at enrollment, gender, religion, and place of residence, were collected through interviews with the patient/caregiver, along with a review of the patient's medical records. Information regarding SCD phenotypes, blood group, and SCD test type administered were extracted from the patient's medical records. Lab results including a complete blood count with differential counts. Follow-up clinical data at each clinic visit were also captured and included vital signs such as body temperature, and blood pressure and management details such as usage of hydroxyurea, penicillin V, malaria chemoprophylaxis, folic acid, up-to-date status of pneumococcal

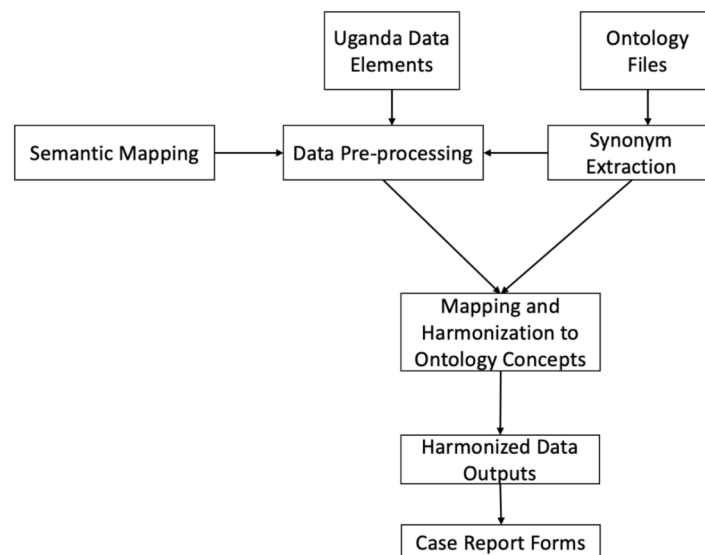


Fig. 1 Figure showing the ontology data mapping and harmonization process

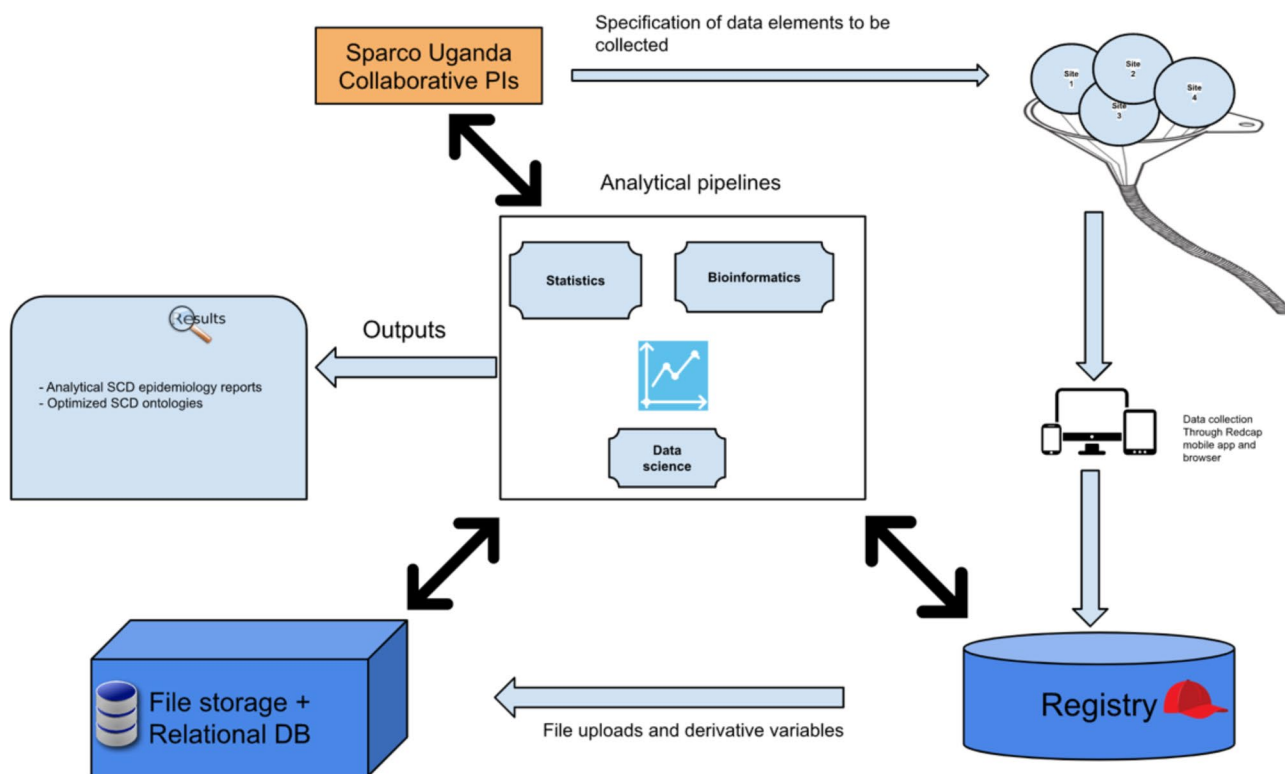


Fig. 2 Figure showing the database design and architecture of the registry

vaccination were captured. Additional clinical information about chest pain, priapism, anemia, and jaundice were also captured as part of the follow-up visit. A deduplication process was implemented within Redcap using unique identifiers to prevent duplicate entries. Enrolling sites had access only to their local patient data ensuring confidentiality and data security. The full data collection instrument with all the variables that are collected has been provided as part of supplementary material (Supplementary Material 1). The data used to generate the figures has been attached in the supplementary materials as two separate files namely enrollment_trends.csv and site-frequencies.csv.

Data management & analysis

Regular data quality checks and progress monitoring were conducted on a weekly basis, accompanied by the timely resolution of any arising queries. The overall data manager conducted comparative analysis and presented aggregated data during these weekly meetings. During the data collection phase, data clerks were afforded ongoing assistance and guidance to navigate and resolve any technical challenges encountered including mobile updates and errors resulting from attaching big sized images. The registry prioritized data quality assurance, employing in-house rules to enforce data type restrictions, range checks, and uniform data entry procedures. In addition to the server’s daily automatic backups, data

was also manually backed up onto an external hard disk system for added security. Furthermore, comprehensive automatic backups were performed quarterly, with data being stored both online and on a distinct external hard disk for redundancy. To ensure further data safety, de-identified data were routinely transmitted to the SPARCo Hub and SADaCC via REDCap, serving as an additional backup layer. A descriptive analysis of the collected data depicting the number of enrollment trends and number of patients at each site was conducted in R 4.2.1 [18], dates were formatted using lubridate v1.8.0 [19] with and the map visualization was done using sf v1.0-15 [20] and rnatuarearth v1.0.1 [21] packages.

Results

From June 2022 to October 2023, we registered and enrolled a total of 5,655 patients with SCD across the 4 sites in Uganda. The geographical distribution of these enrollment sites is depicted in Fig. 3.

Enrollment trends over time

The highest number enrolled in a month was in August 2022(902), and the lowest was in October 2023(56). This is shown in Fig. 4.

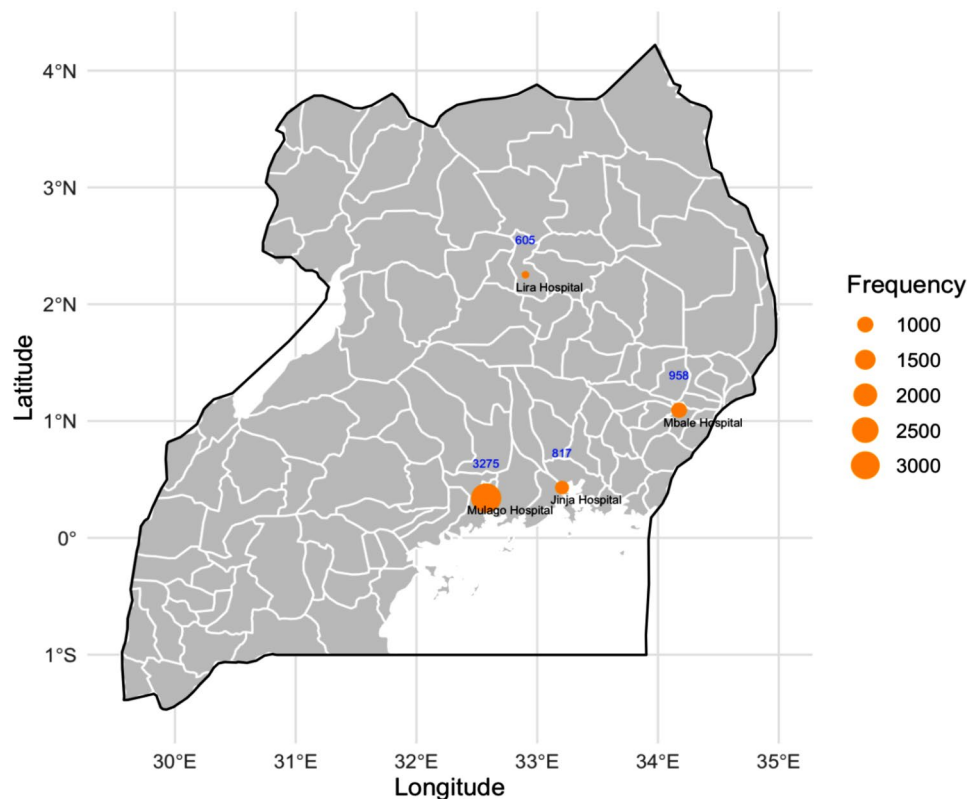


Fig. 3 Map of Uganda showing regions with different sites

Discussion

The SPARCo database has established an extensive registry, distinguished as the inaugural and most extensive of its kind in Uganda. The data components employed by the Ugandan registry mirror those utilized in other nations participating in SPARCo and originate from the SCD ontology, aiming to foster consistency across SCD databases throughout Africa [13, 22]. Although this continent-wide ontology forms the backbone of all registries, national registries can have either more or less ontological elements, depending on their own clinical protocols and standards of care or even simply what data they can afford to collect. In our case, elements from previous forms (Supplementary Material 2) about physical examination, previous diagnosis and whether a patient underwent daycare treatment were added to the ontology.

In collaboration with the SADAaCC, this registry aims for universal acceptance and precise data capture, adhering to the structured SCD ontology. Standard operating procedures (SOPs), devised jointly with SPARCo and SADAaCC, ensure uniform data collection, data quality, and compliance with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, enhancing both human and computer usability [23]. This establishment is similar to SPARCo Ghana, Nigeria and Tanzania registries [24–26]. In contrast with SPARCo Tanzania, the registry enrolled patients from four

regional referral hospitals and district hospitals, with varying SCD epidemiology.

The advancement and coordination of our registry, in collaboration with the broader SPARCo network, are paving the way for the consolidation of data from numerous sites. This consolidation is vital for higher statistical power to drive deeper insights into the landscape of SCD on the continent and improvement of the utilization of the RED-Cap mobile application, particularly for data collection in settings where internet connectivity is unavailable. Previously, the database at Mulago site encountered challenges with lab results often returning a day after they were ordered or some patients lacking funding to order lab tests promptly, leading to inconsistencies in the entered results. However, these issues have been resolved due to the registry's ability to easily update and amend records, ensuring more consistent recording of lab results. The development of harmonized SOPs and minimum uniform data elements across participating sites ensures that the data collected is standardized. This standardization is key in forming a comprehensive dataset within the consortium. Utilizing this extensive data, we are positioned to apply advanced data science approaches, employing innovative methodologies and algorithms designed for analyzing large datasets. Such analysis will be instrumental in yielding critical insights, which are essential for the development of effective interventions and improving the overall care for patients with SCD.

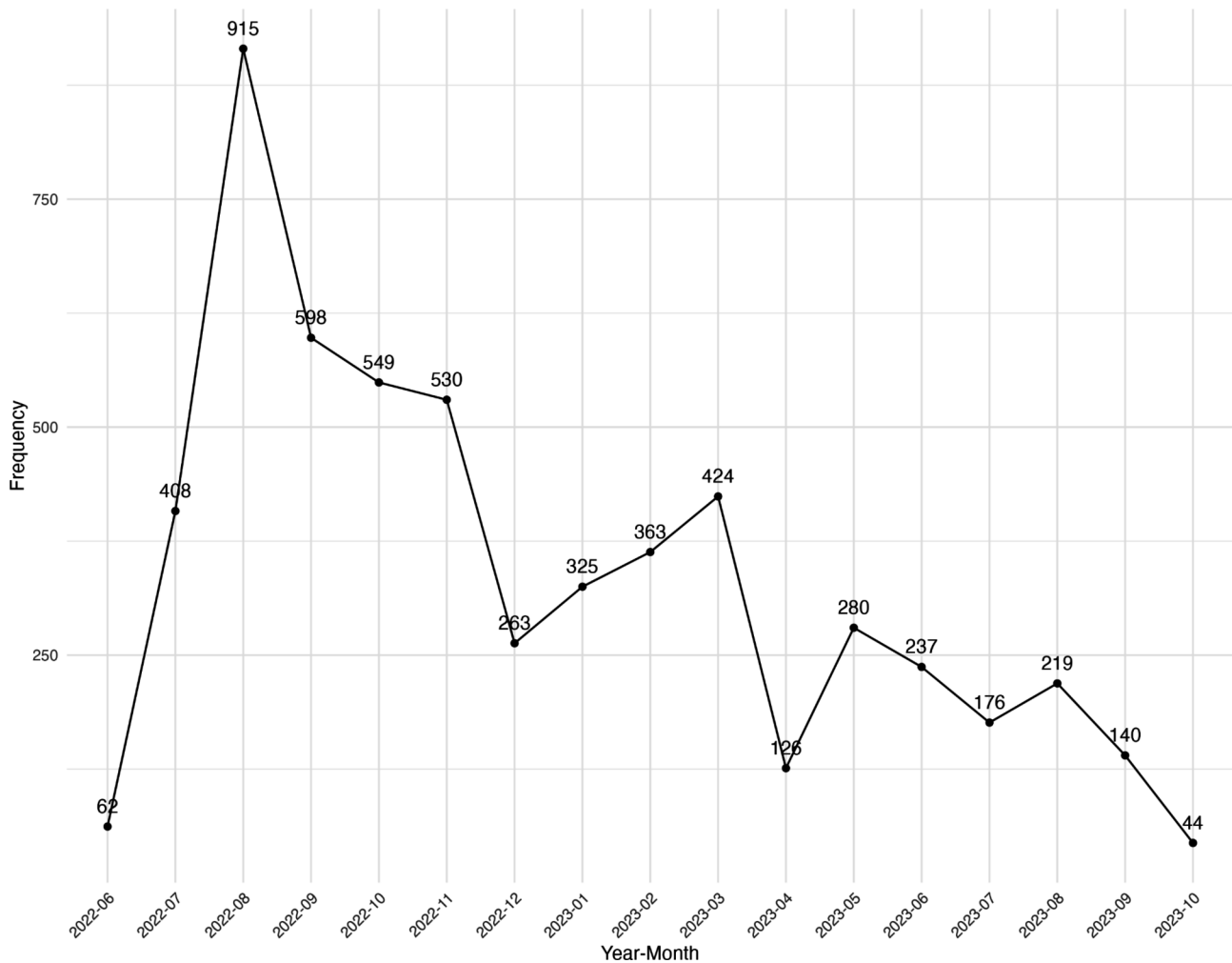


Fig. 4 Enrollment trends of SCD patients from June 2022 to October 2023 showing monthly increases and fluctuations over the study period

The trend in patient enrollment demonstrated a strong and steady increase during the first half-year, with an average of 500 new patients each month. This rate is notably higher than that observed in other SPARCo registries across Africa, such as in Tanzania, Ghana, and Nigeria, where the monthly average was around 150 patients. This highlights the particularly high patient volume in Uganda and the likely more mature SCD clinical infrastructure in the country. Enrollment numbers were highest in August due to the outreach efforts and the addition of two new sites, Jinja and Mbale, which peaked in their first month of enrollment. Initially, only Mulago was operational contributing to the numbers in June and July. The registry experienced a decrease in enrollment in the month of December, this trend has been similarly observed in the other SPARCo sites (Tanzania, Nigeria and Ghana). During December, extended public holidays are observed, prompting many families to celebrate the Christmas festive season away from their regular residences.

Over a period of 17 months, our registry successfully enrolled a total of 5,655 patients, setting a record for the most rapid accumulation of patient data among SPARCo SCD registries within such a timeframe, especially when compared to similar registries in Tanzania, Ghana, and Nigeria [24–26]. The National Referral Hospital, particularly Mulago National Referral Hospital, enrolled more patients than the regional referral hospitals. This discrepancy is attributed to the sickle cell clinic's daily operation, compared to weekly operations at regional hospitals, its larger staff, and its broader service population. The Mulago sickle cell clinic has seen some recent rise in the number of researchers, and patients there might also be more likely to consent. Furthermore, this might be a reflection of increased awareness among the patients on the importance of registries and clinical research. Taken together, the establishment of the registry and the quick accrual of patient numbers in the registry provide evidence of the capacity for sub-Saharan Africa sites to initiate and lead large prospective SCD studies.

Lessons learned

The development of the SCD registry has imparted several valuable lessons, pivotal for the planning and execution of future disease registries. Having robust infrastructure emerged as a cornerstone of our efforts, underscoring the necessity for a solid technological and organizational foundation to support data integrity, management, and scalability. The infrastructure, which was provided by the NIH-funded African Center of Excellence in Bioinformatics and Data Intensive Sciences, has also supported various technological and infrastructural initiatives, further enhancing research and development capabilities across the continent [17, 27–29]. This was complemented by the formation of a multidisciplinary team, bringing together clinicians, software developers, data managers, and policymakers. This diverse collaboration not only facilitated a comprehensive approach to registry development but also ensured that the system was responsive to the needs of all stakeholders, thereby enhancing the utility and applicability of the registry in various research and clinical contexts. Furthermore, leveraging REDcap, an existing database platform – and only making modifications and customizations to its core relational architecture to fit local context, did not only save development resources but also quickened interoperability with other national registries within the consortium.

The approach of starting small and then expanding proved to be instrumental in the registry's development, allowing for the initial focus on a limited set of data elements and sites. This methodology enabled the team to identify and address issues in a controlled environment, refining data collection processes before scaling up. Such a phased expansion was crucial for maintaining system stability and adaptability, ensuring that the registry could accommodate a growing scope of data and a wider network of participating sites without compromising on data quality or system performance.

The implementation of extensive training programs for all involved parties, especially data entrants, was another critical factor in the registry's success. Ensuring that individuals were well-versed in the registry's operations and objectives significantly contributed to the accuracy and reliability of the data collected. Moreover, achieving comprehensive geographical coverage ensured the collection of data that was representative of the entire country, capturing the diversity of SCD manifestations and treatments. This wide-ranging data collection was further enhanced by the inclusion of local language support, making the registry more accessible and inclusive, thus broadening the spectrum of participants and enriching the dataset. Connectivity issues in remote areas hindered consistent data collection, a challenge that was addressed through the adoption of mobile data collection tools that allowed for offline data capture.

Conclusion

The establishment of the SPARCo Uganda registry represents a significant milestone in the realm of SCD research, particularly within the context of resource-limited environments. This endeavor demonstrates the viability of developing comprehensive disease registries in such environments and underscores the imperative of integrating individuals diagnosed with SCD seamlessly into the healthcare system, ensuring they receive the specialized care they need.

The findings from the SPARCo Uganda registry, which are currently being used to facilitate a range of studies, including the analysis of SCD clinical phenotypes, newborn screening, and evaluation of hydroxyurea use, offer a promising direction for SCD management, research, and data-driven approaches in Uganda. They suggest that implementing a nationwide electronic registry for hemoglobinopathies, with patient consent, could revolutionize SCD research and patient care.

Abbreviations

ACE	African Center of Excellence in Bioinformatics and Data sciences
CRF	Case Report Form
SCD	Sickle cell disease
SPARCo	Sickle Pan Africa Research Consortium
SADaCC	Sickle Africa Data Coordinating Center

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-024-02618-9>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

We would like to acknowledge Isaac Birungi, Aisha Nakirya, Nathan Bikala, Jacob Isabirye, Geoffrey Ssebagala, Tamuzadde Paul, Stella Nalubega, the data entrants at the four hospitals. The authors would also like to acknowledge the NIH-funded African Center of Excellence in Bioinformatics and Data Science facilities based at the Infectious Diseases Institute at Makerere University who provided the computational infrastructure.

Author contributions

Authors' contribution MN: Data analysis, Data Interpretation, Writing-Original draft; HM: Data analysis, Writing-Final draft; DM: Data Interpretation, Writing-Original draft; DJ: Data analysis, Data Interpretation, Writing-Final draft; RN: Data Interpretation, Writing-Original draft; RO: Conception; PK: Conception; GN: Data Interpretation; Writing-Final draft; GK: Writing-Final draft; EM: Conception; IB: Data Management; HH: Conception; JM: Data Management; MJ: Data Management; VN: Conception; AW: Conception; JM: Conception; SK: Conception, Data Interpretation, Writing-Final draft. All authors read and approved the final manuscript.

Funding

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number U01HL156914. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

The ethical approval to conduct this study was obtained from Makerere University School of Medicine Research and Ethics Committee (SOMREC) under number Mak-SOMREC-2021-170 and the Uganda National Council of Science and Technology under number HS2012ES. We obtained administrative clearance from all the 4 hospitals to set up the registries. Written informed consent was acquired from each participant, and for patients under 18 years old, their caregivers provided the informed consent. The study was performed in accordance with the ethical standards in the 1964 Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Health Sciences, University of Bristol, Bristol BS40 5DU, UK

²Jean Golding Institute, University of Bristol, Bristol BS8 1UH, UK

³The Infectious Diseases Institute, Makerere University, P. O. Box 22418, Kampala, Uganda

⁴The African Center of Excellence in Bioinformatics and Data Intensive Sciences, Kampala, Uganda

⁵Department of Computer Science, College of Computing and Information Sciences, Makerere University, P. O. Box 7062, Kampala, Uganda

⁶Department of Pediatrics and Child Health, School of Medicine, Makerere University, Kampala, Uganda

⁷Undergraduate Medical Education, Aga Khan University Medical College, Nairobi, Kenya

⁸Department of Pediatrics, Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Canada

⁹Global Health Uganda, Kampala, Uganda

¹⁰Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

¹¹Department of Genetic Medicine, McKusick-Nathans Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

¹²Department of Haematology and Blood Transfusion/Sickle Cell Program, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

Received: 4 June 2024 / Accepted: 23 July 2024

Published online: 29 July 2024

References

- Piel FB, Rees DC, DeBaun MR, Nnodu O, Ranque B, Thompson AA, et al. Defining global strategies to improve outcomes in sickle cell disease: a lancet haematology commission. *Lancet Haematol*. 2023;10:e633–86.
- Weatherall DJ. Genetic variation and susceptibility to infection: the red cell and malaria. *Br J Haematol*. 2008;141:276–86.
- Shriner D, Rotimi CN. Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the holocene wet phase. *Am J Hum Genet*. 2018;102:547–56.
- Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *Lancet Lond Engl*. 2010;376:2018–31.
- Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M, et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geo-statistical model-based map and population estimates. *Lancet Lond Engl*. 2013;381:142–51.
- Ndeezi G, Kiyaga C, Hernandez AG, Munube D, Howard TA, Ssewanyana I, et al. Burden of sickle cell trait and disease in the Uganda sickle surveillance study (US3): a cross-sectional study. *Lancet Glob Health*. 2016;4:e195–200.
- Wonkam A, Makani J. Sickle cell disease in Africa: an urgent need for longitudinal cohort studies. *Lancet Glob Health*. 2019;7:e1310–1.
- Makani J, Cox SE, Soka D, Komba AN, Oruo J, Mwamtemi H, et al. Mortality in sickle cell anemia in Africa: a prospective cohort study in Tanzania. *PLoS ONE*. 2011;6:e14699.
- Wonkam A, Ponde C, Nicholson N, Fieggen K, Ramessar R, Davidson A. The burden of sickle cell disease in Cape Town. *S Afr Med J*. 2012;102:752–4.
- Diallo DA, Guindo A. Sickle cell disease in sub-saharan Africa: stakes and strategies for control of the disease. *Curr Opin Hematol*. 2014;21:210–4.
- Makani J, Sangeda RZ, Nnodu O, Nembaware V, Osei-Akoto A, Painsil V, et al. SickleInAfrica. *Lancet Haematol*. 2020;7:e98–9.
- UBOS. National population and housing census 2014 Main Report. 2014. <https://web.archive.org/web/20171012091133/http://www.ubos.org/online-files/uploads/ubos/NPHC/2014%2520National%2520Census%2520Main%2520Report.pdf>. Accessed 17 Mar 2024.
- Mulder N, Nembaware V, Adekile A, Anie KA, Inusa B, Brown B et al. Proceedings of a sickle cell disease ontology workshop — towards the first comprehensive ontology for sickle cell disease. *Appl Transl Genomics*. 2016;9:23–9.
- Sickle Cell Disease Ontology Working Group. The sickle cell disease ontology: enabling universal sickle cell-based knowledge representation. *Database*. 2019;2019:baz118.
- Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inf*. 2019;95:103208.
- Pasalic D, Reddy JP, Edwards T, Pan HY, Smith BD. Implementing an electronic data capture system to improve clinical workflow in a large academic radiation oncology practice. *JCO Clin Cancer Inf*. 2018;1–12.
- Giovanni MY, Whalen C, Hurt DE, Ware-Allen L, Noble K, McCarthy M, et al. African centers of excellence in bioinformatics and data intensive science: building capacity for enhancing data intensive infectious diseases research in Africa. *J Infect Dis Microbiol*. 2023;1:006.
- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
- Grolemund G, Wickham H. Dates and times made easy with lubridate. *J Stat Softw*. 2011;40:1–25.
- Pebesma E. Simple features for R: standardized support for spatial vector data. *R J*. 2018;10:439–46.
- Massicotte P, South A. rnaturalearth: world map data from natural earth. 2023.
- Nembaware V, Mazandu GK, Hotchkiss J, Safari Serufuri J-M, Kent J, Kengne AP, et al. The sickle cell disease ontology: enabling collaborative research and co-designing of new planetary health applications. *Omic J Integr Biol*. 2020;24:559–67.
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
- Kandonga D, Sangeda RZ, Masamu U, Kazumali E, Jonathan A, Msangawale M et al. Development of the sickle pan-african research consortium registry in Tanzania: opportunity to harness data science for sickle cell disease. *Front Hematol*. 2023;2.
- Nnodu O, Madu A, Chianumba R, Isa HA, Olanrewaju I, Osagie S et al. Establishing a database for sickle cell disease patient mapping and survival tracking: the sickle pan-african research consortium Nigeria example. *Front Genet*. 2022;13.
- Painsil V, Amuzu EX, Nyanor I, Asafo-Adjei E, Mohammed AR, Yawnumah SA et al. Establishing a sickle cell disease registry in Africa: experience from the sickle Pan-African research consortium, Kumasi-Ghana. *Front Genet*. 2022;13.
- Nsubuga M, Galiwango R, Jjingo D, Mboowa G. Generalizability of machine learning in predicting antimicrobial resistance in *E. Coli*: a multi-country case study in Africa. *BMC Genomics*. 2024;25:287.
- Please H, Narang K, Bolton W, Nsubuga M, Luweesi H, Richards NB et al. Virtual reality technology for surgical learning: qualitative outcomes of the first virtual reality training course for emergency and essential surgery delivered by a UK–Uganda partnership. *BMJ Open Qual*. 2024;13.
- Buyego P, Katwesigye E, Kebirungi G, Nsubuga M, Nakyejwe S, Cruz P, et al. Feasibility of virtual reality based training for optimising COVID-19 case handling in Uganda. *BMC Med Educ*. 2022;22:274.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.