DATA SCIENCE JOURNAL

Kumuthini, J, et al. 2019. Proposed Guideline for Minimum Information Stroke Research and Clinical Data Reporting. *Data Science Journal,* 18: 26, pp. 1–18. DOI: https://doi.org/10.5334/dsj-2019-026

RESEARCH PAPER

Proposed Guideline for Minimum Information Stroke Research and Clinical Data Reporting

Judit Kumuthini[1], Lyndon Zass[1], Melek Chaouch[2], Michael Thompson[3], Paul Olowoyo[4], Mamana Mbiyavanga[5], Faniyan Moyinoluwalogo[6], Gordon Wells[1], Victoria Nembaware[5], Nicola J. Mulder[5], Mayowa Owolabi[6], H3ABioNet Consortium's Data and Standards Working Group as members of the H3Africa Consortium

[1] Centre for Proteomic and Genomic Research, Cape Town, ZA

[2] Laboratory of Bioinformatics, Biomathematics and Biostatistics, Institut Pasteur de Tunis, Tunis, TN

[3] National Institute of Mathematical Sciences, Kumasi, GH

[4] Federal Teaching Hospital, Ido-Ekiti/Afe Babalola University, Ado-Ekiti, NG

[5] Computational Biology Division, IDM, University of Cape Town, Cape Town, ZA

[6] University of Ibadan, Ibadan, NG

Corresponding author: Dr. Judit Kumuthini (Judit.kumuthini@cpgr.org.za)

The management and analyses of large datasets is one of the grand challenges of modern biomedical research. Establishing methods to harmonise and standardise data collection, reporting, sharing and the employed data dictionaries, can support the resolution of these challenges whilst improving research quality, data quality and integrity, allowing sustainable knowledge transfer through re-usability, interoperability, reproducibility. The current project aimed to develop and propose a standardised reporting guideline for stroke research and clinical data reporting. Through systematic consolidation and harmonization of published data collection and reporting standards, several recommendations were drafted for the proposed guideline. These recommendations were reviewed by domain-researchers and clinicians using an online survey, developed in REDCap. The survey was completed by 20 international stroke-specialists, majority of respondents were based in Africa (10), followed by America, Europe and Australia (10). Of these respondents; the majority were working as dual clinician-researchers (57%) with more than 10 years' experience in the field (78%). Data elements within the reporting standard were classified as participant-, study- and experiment-level information, further subdivided into essential or optional information, and defined using existing ontologies. The proposed reporting guideline can be employed for research utility and adapted for clinical utility as well. It is accompanied with an associated XML schema for REDCap implementation, to increase the user friendliness of data capturing, sharing, reporting and governance. Ultimately, the adoption of common reporting in stroke research has the potential to ensure that researchers gain the maximum benefit from their generated data and data collections.

Keywords: Stroke; minimum information requirement guideline; standardization; reporting guideline; data reporting; H3ABioNet

Introduction

High-throughput technologies are increasingly being employed in biomedical- and healthcare-informatics research, producing large biological and clinical data sets at rapid speed (Luo et al., 2016). Modern biomedical and clinical research is thus characterised by the exponentially increasing volume of a variety of data types and structures, produced and processed at unprecedented velocity (Luo et al., 2016). Integrating these different sources of information holds great potential to elucidate the aetiologies of complex medical conditions, develop novel treatments for such conditions and revolutionise modern health care with the incorporation

of personalised medicine, predictive modelling and clinical decision support, and improved disease and safety surveillance (Lee and Yoon, 2017).

Stroke, defined as an acute focal or global neurological deficit, results from spontaneous haemorrhage or infarction of the central nervous system with objective evidence of infarction or haemorrhage irrespective of duration of clinical symptoms (Sacco et al., 2013). It remains one of the primary causes of brain injury, disability and death, worldwide (Benjamin et al., 2017; Wang et al., 2016). Currently, big data analytics are employed in stroke research to elucidate the genetic and environmental underpinnings of stroke (Akinyemi et al., 2015; Owolabi et al., 2018), as well as to research improved methods of stroke health care, such as revolutionising visual analytics, employing predictive analytics in hypertension patients and employing telecardiology as method of care (Wang and Alexander, 2016).

Large-scale data analytics have introduced new challenges to biomedical researchers, including the storage, management, sharing and analyses of datasets (Luo et al., 2016). The datasets require powerful and novel technologies to extract biologically meaningful information and conclusions, as well as enable more broad-based health-care solutions (Luo et al., 2016). Standardising the methods in which clinical and research data are collected, reported, shared, managed and (or) stored can support the resolution of these challenges, by enhancing data compatibility, interoperability, reproducibility and re-use (Skrocki, 2013), and facilitating data sharing and collaboration (Khan, 2017). This can be particularly useful for stroke research in low-resource settings, where primary research is historically epidemiological in nature, and the technological capacity for appropriate case identification is lacking (Adeloye, 2014).

Biomedical standardisation efforts are being led globally by the Global Alliance for Genomics and Health (GA4GH) (www.ga4gh.org), a policy-framing and technical standards-setting organization, as well as FAIRsharing (www.fairsharing.org), a dynamic standards database which aims to promote FAIR (Findable, Accessible, Interoperable, and Reusable) principals (Wilkinson et al., 2016). Additionally, the PhenX (consensus measures for Phenotypes and eXposures) Toolkit (https://original-phenxtoolkit.rti.org/index.php) promotes standardised data collection, by recommending a catalogue of standard measures of phenotypes and environmental exposures for use in biomedical research, although some measures are limited in terms of applicability to low-resource settings (Hamilton et al., 2011).

Drawing from the aforementioned resources, as well as existing data collection measures, data reporting standards, controlled vocabularies, data dictionaries and relevant ontologies, the Human Heredity and Health in Africa's (H3Africa) Bioinformatics Network's (H3ABioNet, www.h3abionet.org) (H3Africa Consortium et al., 2014; Mulder et al., 2016) Data & Standards work package aimed to develop domain/field-specific research data reporting guidelines for several diseases. The current report focused on the development of the stroke-specific research data reporting guideline entitled, 'The Minimum Information Required Guideline for Stroke Research and Clinical Data Reporting (Version 1.0)'.

Methods

A list of reporting recommendations for the guideline was proposed based on the review of existing literature and online resources. These resources included data collection methods hosted on PhenX Toolkit, including the measure for collecting stroke history, and experimental reporting guidelines, hosted on FAIRsharing, including The Minimum Information About a Proteomics Experiment (MIAPE) (Taylor et al., 2007) and The Minimum Information required for DMET experiment (MIDE) (Kumuthini et al., 2016). These recommendations were also harmonised with the H3Africa Standard Case Report Form (CRF) (www.h3abionet.org/data-standards/datastds). A reporting guideline was drafted based on these recommendations, which was then subdivided into three main sections, entitled; participant-level information (which details information specific to the study participants), study-level information (which details information specific to the study) and experiment-level information (which details information specific to the experiments within a study), to illustrate the different levels and users of the data.

Thereafter, the drafted guideline was reviewed by a broad range of international stroke researchers and clinicians, using an anonymous online survey, constructed to evaluate, harmonise and consolidate the proposed recommendations, identify which recommendations represented "essential" or "optional" information, propose additional recommendations and remove any existing reporting, ontology, nomenclature, and unit inconsistencies. The online survey was constructed, and study data were collected and managed using REDCap (Research Electronic Data Capture) 7.5.0, hosted at The Centre for Proteomic and Genomic Research (CPGR) (Harris et al., 2009). The online survey consisted of 116 fields (Supplementary File 1).

Once harmonised, the recommendations (henceforth referred to as elements) were manually defined using ontologies found through the BioPortal search engine (Noy et al., 2009), the Ontology Lookup Service (OLS) at the European Bioinformatics Institute (EMBL-EBI) (Côté et al., 2006) and the Zooma annotation tool.

Following completion, an associated eXtensible Markup Language (XML) schema was designed for implementation in REDCap. The schema was designed to carry all the data and metadata within the reporting guideline, while maintaining the associations between them to allow the exchange of clinical data between dissimilar health information or research systems without losing the semantics and structure of the reported data. The schema also defines the rules of validation for each element, as well as the datatype, atomic units and validation rules for each element, to ensure reporting correctness.

Results

The online survey was completed by 20 international stroke-specialists, majority of respondents were based in Africa (10), followed by America (4), Europe (4) and Australia (2). Of these respondents; 28% were working as clinicians, 15% were working as researchers, and 57% were working as dual clinician-researchers. The majority of respondents had between 10- and 20-years' experience in the field (38%), whilst 34% of respondents had more than 20 years' experience in the field, 15% had between 5- and 10-years' experience in the field and 13% had less than 5 years' experience in the field. **Figures 1** and **2** illustrate the survey response to the proposed elements. Furthermore, respondents proposed additional elements, which shaped the final structure of the reporting guideline. This includes, but is not limited to, Diet and Dyslipidaemia. The raw survey results can be found in Supplementary File 2.

The Minimum Information Required Guideline: Stroke Research and Clinical Data Reporting is summarised in **Table 1**.

The quintessential information reported using the standard are separated into three fields; participant-, study- and experiment-level information. The standard further divides elements into essential and optional information. Optional elements refer to information which is not necessary for the interoperation of studies within the same field but useful for integrating studies from varying disease fields. Participant-level information contains 13 subsections of varying essential and optional elements, including Demographics, Lifestyle Factors, Anthropometrics, Blood Pressure, Adverse Drug Reactions, Urine-Related Test Index, Stroke History,

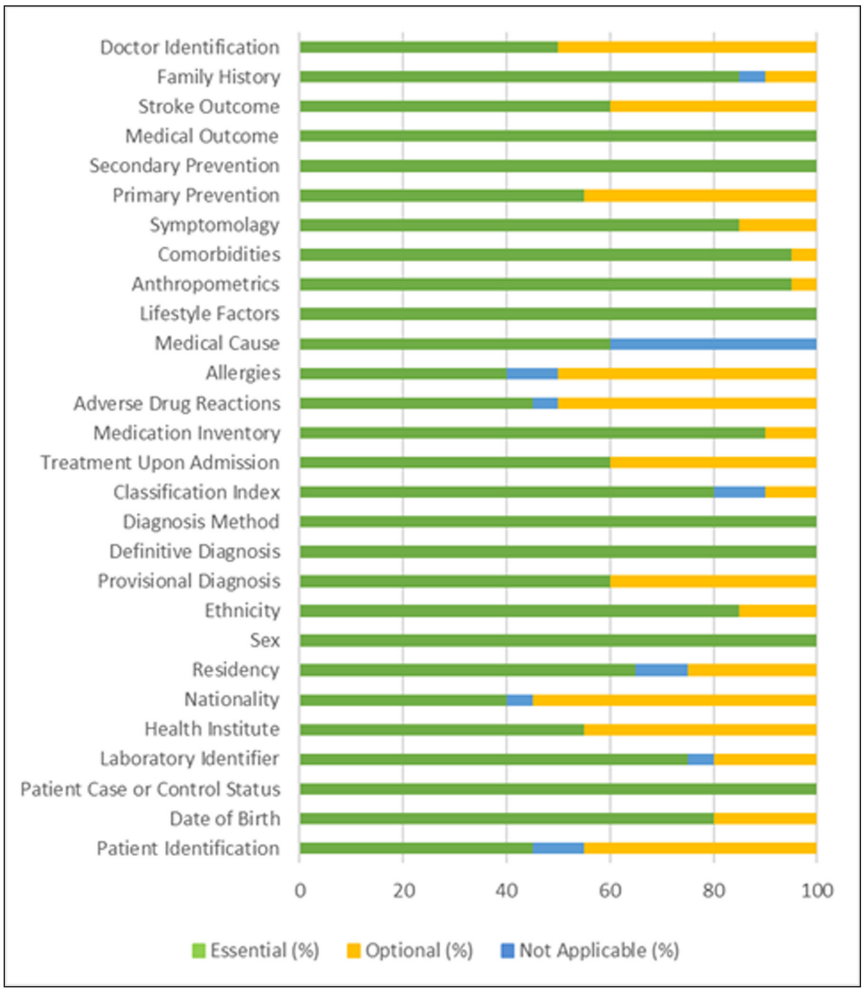

Figure 1: Survey response to proposed participant-level information.

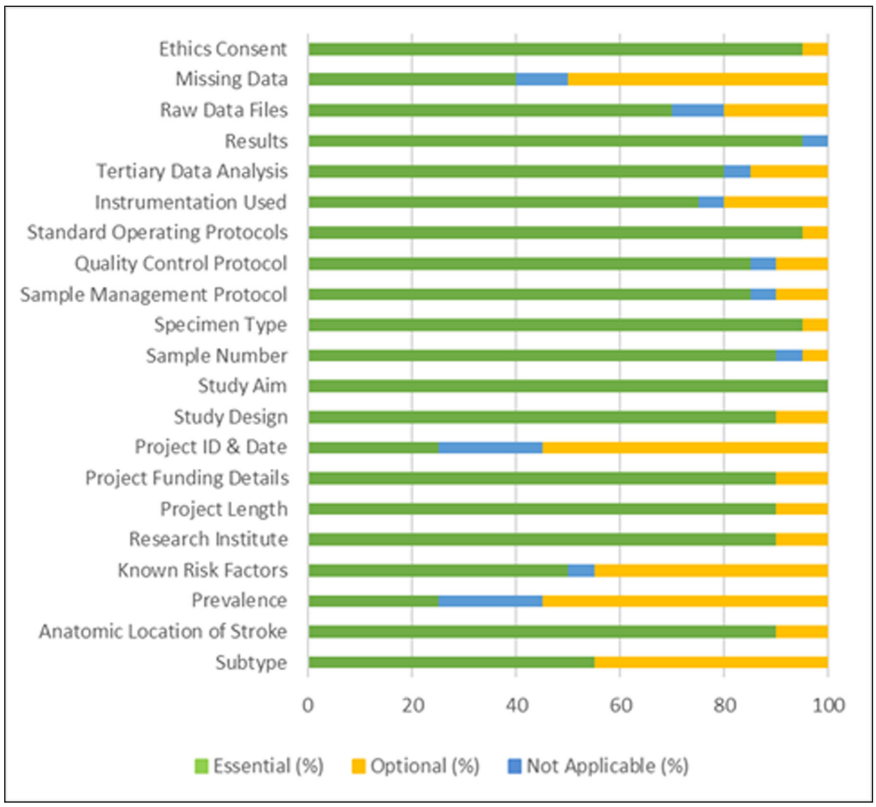

Figure 2: Survey response to proposed study- and experiment-level information.

Sample-Specific Information, Stroke-related Information, Prescribed Medication, Non-Prescribed Medication, and Therapy. The Study-level information includes various elements which describe the details of a given study, including essential elements such as Study ID, Research Institute and Study Design, and optional elements such as Study Duration, Study Start Date, and Pubmed Unique Identifier. Finally, experiment-level information includes various elements which describe the various experiments within a given study, including essential elements such as Biospecimen Type, Instrumentation employed, Sample Management Protocol, Quality Control Protocol and Experimental Aim, and optional elements such as Output Location, which describes where the data will be saved. Although descriptions of the Output Location are widely encouraged, this data element remains optional to accommodate scenarios where data is private, under embargo and(or) the reporting guideline is used explicitly for internal research data management.

The complete reporting guideline can be found in Supplementary File 3, specifying each element's data type, collection format and (or) accepted values, and related ontologies and standards. Herein, the Ontology ID column contains the most appropriate ontology which the element is mapped to whilst the Concordant Ontologies and Concordant Standards columns describe ontologies and standards which include similar data elements. These lists are not meant to be comprehensive or exhaustive, but to illustrate the utilization and overlap with existing resources. A comprehensive guideline explaining how to employ the reporting guideline locally can also be found in Supplementary File 4.

An associated XML schema was developed for REDCap implementation, consisting of 3 sections – the participant-, experiment and study-level information, and can be found in Supplementary File 5. The relationship between these sections are illustrated in Supplementary File 6. The XML schema represents and requests information as outlined in the proposed reporting guideline, and therefore functions as a standard format of the reporting guideline. Importantly, the reporting guideline and the associated XML schema can also be obtained from the H3ABioNet website (www.h3abionet.org/data-standards/datastds), along with a guideline document on how to employ the reporting guideline locally.

Discussion

The paper outlines the development of the Minimum Information Required: Stroke Research and Clinical Data Reporting Guideline. To our knowledge, though ontologies and collection standards have previously been described for stroke-related clinical care and research, no reporting guideline for stroke research and

Table 1: The Minimum Information Required Guideline: Stroke Research and Clinical Data Reporting.

Elements			Importance	Definition
Participant-level information				
Demographics	Date of Birth		E	The calendar date on which a participant was born.
	Sex		E	The classification of the participant's sex.
	Self-reported Race/Ethnicity		E	Membership to social group based on a common heritage.
	Country of Birth (COB)		E	The country that the participant was born in.
	Country of Residence		O	The country that the participant resides in.
	Native Language(s)		E	The primary systematic means of communication used in the participant's household.
	Tribal Affiliation		O	The tribe which the participant is affiliated to.
	Father's Country of Birth		O	The country in which the participant's biological father was born.
	Mother's Country of Birth		O	The country in which the participant's biological mother was born.
Lifestyle Factors	History of Hypertension	Has a healthcare worker ever said that you have high blood pressure or hypertension?	E	Participant's background regarding high blood pressure or hypertension.
		If yes, then at what age were you first told this?		
		FOR WOMEN ONLY: Was this during pregnancy?		
		Have you ever taken medication for hypertension/high blood pressure?		
		If yes, then at what age did you begin taking medicine for this?		
	Physical Activity	7-Day Frequency	E	The number of occurrences of physical activity per unit time (7 days).
		Time		The average time spent per physical activity (in min).

(Contd.)

Elements			Importance	Definition
		Intensity		The average energy expended per physical activity. Light exercise is 20–60 minutes and elevates heart rate to 35–60% of maximum heart rate (e.g. housework, gardening, slow walking); moderate exercise is 20–60 minutes and elevates heart rate to 35–60% of maximum heart rate (e.g. basketball, single tennis, brisk walking); strenuous exercise elevates heart rate to over 60% of maximum heart rate (e.g. jogging, swimming, bicycling).
	Alcohol Use	Lifetime Use	E	A description of an individual's current and past experience with alcoholic beverage consumption.
		Age of Initiation		The age of initiation of alcoholic beverage consumption.
		30-Day Frequency		The number of occurrences of alcoholic beverage consumption per unit time (past 30 days).
		30-Day Quantity		A record of the quantity of alcohol consumption (in standard drinks) (past 30 days).
	Tobacco Use	Lifetime Use	E	Record of whether the participant has ever used any tobacco product during his or her entire life.
		Lifetime Frequency		
		Age of Initiation		The age of initiation of tobacco use.
	Recreational Drug Use	Lifetime Use	E	Record of whether the participant has ever used a drug during his or her entire life.
		Age of Initiation		The age of initiation of drug use.
		30-Day Type		A record of the participant's type of drug use within the past 30 days.
		30-Day Frequency		The number of occurrences of drug use per unit time (past 30 days).
	Diet		E	The customary allowance of food and drink taken by a person from day to day.
Anthropo-metrics	Average Height		E	The vertical measurement of distance from the sole to the crown of the head with body standing on a flat surface and fully extended (in cm). Averaged over 3 measurements.

(Contd.)

Elements			Importance	Definition
	Average Weight		E	The measurement of mass or quantity of heaviness of an individual (in kg). Averaged over 3 measurements.
	Waist Circumference		O	The abdominal circumference at the navel (in cm).
	Head Circumference		O	A circumferential measurement of the head at the widest point (in cm).
	Body Surface Area		O	A measure of the 2-dimensional extent of the body surface (i.e. the skin) (in m^2).
	Prosthesis (if applicable)		E	Location of a device which is an artificial substitute for a missing body part or function.
Blood Pressure	Average Systolic Blood Pressure		E	The average pressure exerted into the systemic arterial circulation during the contraction of the left ventricle of the heart. (in mmHg).
	Average Diastolic Blood Pressure			The average pressure exerted into the systemic arterial circulation during cardiac ventricular relaxation and filling (in mmHg).
Adverse Drug Reactions	Medication		O	The drug product which caused a detrimental or unintended response associated with the use of a medication.
	Type			The type of detrimental or unintended response associated with the use of a medication.
	Date			The calendar date on which the ADR occurred.
Dyslipidaemia	High-Density Lipoprotein (HDL)		E	A lipoprotein metabolism disorder characterized by decreased levels of high-density lipoproteins, or elevated levels of plasma cholesterol, low-density lipoproteins and/or triglycerides.
	Low-Density Lipoprotein (LDL)			
	Triglycerides			
Stroke History	History of Stroke	Were you ever told by a doctor or healthcare worker that you had a stroke, TIA, mini-stroke, transient-ischemic attack?	E	A question to determine if the respondent has had a stroke and/or any symptoms related to this event.
		Have you ever had a sudden painless weakness or numbness on one side of the body, suddenly lost one half of your vision, lost the ability to understand what		

(Contd.)

Elements			Importance	Definition
		people are saying or lost the ability to express yourself verbally or in writing?		
	Family History of Stroke	Has anyone in your family had a stroke?	E	A record of a patient's background regarding stroke and stroke-related events of blood relatives.
	Primary Prevention		E	Primary prevention involves prevention of disease in susceptible individuals or populations through promotion of health and specific protection as distinguished from the prevention of complications or after-effects of existing disease.
	Secondary Prevention		E	Secondary prevention involves procedures or treatment processes designed to prevent further complications.
	Consanguinity	Any cases of consanguineous mating in the family? 'If Yes, please specify:'	O	Reproduction between genetically related individuals.
Sample-specific Information	Sample Identifier		E	Name or other identifier of an entry from a biosample database.
	Sample's Case or Control Status		O	An indication of a subject's status as a case or a control for a given study.
	Consent		E	The planned process that an individual agrees to participate in.
Stroke-related Information	Differential Diagnosis		E	Differential diagnosis refers to the process of differentiating one diagnosis from another, and in turn, providing the most fitting diagnosis based on an individual's presentation.
	Date of Diagnosis		O	The calendar date of confirmatory diagnostic testing.
	Stroke Scale		O	Classification systems employed for clinical and research purposes which to improve diagnostic accuracy, determine the suitability of specific treatments, monitor change in neurologic impairments, and predict and measure outcomes.
	Instrumentation		O	The specialized objects, or items of electrical or electronic equipment, employed to perform diagnosis (with versions).
	Clinical Signs		E	The objective evidence of disease perceptible to the examining healthcare worker.
	Stroke Outcome		E	The result of an action (stroke).

(Contd.)

Elements		Importance	Definition
	Stroke Impact	E	Disabilities and impairments due to a stroke.
	Histopathology	E	The visual examination of cells or tissue (or images of them) with an assessment regarding the quality of the cells or tissue.
	Pre-Stroke Co-morbidities (Systemic)	E	The presence of co-existing or additional medical conditions pre-stroke.
	Post-Stroke Co-morbidities (Systemic)	E	The presence of co-existing or additional medical conditions post-stroke.
	Pathogenic Co-morbidities	E	The presence of co-existing or additional pathogenic diseases with reference to an initial diagnosis or with reference to the index condition that is the subject of study.
	Allergies	O	An immune response or reaction to substances that are usually not harmful.
Prescribed Medication	Medication	E	A record of the prescribed drug product currently in use.
	Dosage		The size or frequency of a dose of a medicine or drug.
	Strength		The amount of the medicine or drug that provides its particular effect.
	Reason		The cause of the prescription.
	Start Date		The calendar date on which treatment was initiated.
	Stop Date		The calendar date on which treatment is to be or was terminated.
Non-Prescribed Medication	Medication	E	A record of the non-prescribed drug product use in the past 2 weeks.
	Dosage		The size or frequency of a dose of a medicine or drug.
	Reason		The cause of the prescription.
	Start Date		The calendar date on which treatment was initiated.
	Stop Date		The calendar date on which treatment is to be or was terminated.

(Contd.)

Elements		Importance	Definition
Study-level information			
Study-specific information	Research Institute	E	The name of the organisation affiliated with a specific study.
	Study Duration	O	The duration of any specifically defined piece of work that is undertaken or attempted to meet requirements (in years).
	Study Start Date	O	The calendar date on which the project is initiated.
	Study ID	E	The unique identifier of the project.
	Disease	E	Clinical entity defined by a set of phenotypic abnormalities resulting from a common physiopathological mechanism with a homogeneous evolution and homogeneous therapeutic possibilities.
	Clinical Subtype	O	The subdivision of a disease, malformation syndrome, morphological anomaly, biological anomaly, clinical syndrome or particular clinical situation in a disease or a syndrome further defined by its particular clinical presentation.
	Study Design	E	The nature of the investigation or the investigational use for which clinical study is being done.
	Study Aim	E	A textual entity describing the study aim.
	Sample Size	E	The subset number of a larger population, selected for investigation to draw conclusions or make estimates about the larger population.
	PMID	O	PubMed unique identifier of an article.
	DOI	O	Digital Object Identifier (DOI) of a published article.
Experiment-level information			
General	Biospecimen Type	E	The type of a material sample taken from a biological entity for research purposes.
	Sample Management Protocol	E	The specifications employed for the management of samples.
	Quality Control Protocol	E	The specifications employed to ensure a certain level of quality of biospecimens.

(Contd.)

Elements		Importance	Definition
	Experimental Aim	E	A textual entity describing the experimental aim.
	Experimental Protocol	E	The specifications with respect to the design and implementation of an experiment or set of experiments.
	Instrumentation	E	Specialised equipment, tools, appliances, and(or) apparatus employed in the experiment(s).
	Data Analysis	E	The data transformation techniques used to analyse and interpret the data to gain a better understanding of it.
	Experimental Result	E	The outcome of the experiment or set of experiments.
	Output Location	O	Full name and location of output (raw or analysed data).

Footnote: E – Essential; O – Optional.

clinical data has previously been proposed or published. Most notably, the Stroke Ontology (https://bioportal.bioontology.org/ontologies/STO-DRAFT) defines the terms and relationships of the knowledge domain of stroke and The Human Phenotype Ontology (HPO) (Robinson and Mundlos, 2010) defines stroke-related phenotypes. Similarly, PhenX, Clinical Data Interchange Standards Consortium (CDISC) and Health Level Seven (HL7) have previously developed and proposed standards fit for clinical data collection in various disease fields. In the development of our reporting guideline, we utilised these existing resources to harmonise collection measures and terms and develop a comprehensive and harmonised data management tool which allows centralised management of both clinical and research data in a complex disease field (stroke) which requires collaborative and inter-disciplinary research. Combining the clinical and research data elements in one standard allows principal investigators to maintain various levels of data access whilst still centralising comprehensive data management and storage. This empowers principal investigators to manage their research data in a coordinated and comprehensive manner, and to maintain the participant-level data associated with various studies and(or) experiments in a user-friendly way (and vice versa).

Employing the reporting guideline can thus add great benefit to stroke research studies, as it references stroke-based ontologies, data dictionaries and collection standards, ensuring comprehensive, harmonised data reporting, which is re-usable and enhances interoperability. The reporting guideline is designed for use by research clinicians and healthcare workers, researchers, data managers and bioinformaticians involved in stroke research, bearing in mind different levels of data access. Given the appropriate levels of data input and access right, allows the reporting guideline to be used in both research and clinical settings whilst defining the information as essential or optional allows the research to be adaptable for various types of research with regards to stroke. Additionally, the reporting guideline goes beyond listing "minimum required" data elements and aims to provide a comprehensive data dictionary and controlled vocabulary with standardised response options, which is scalable and can be adapted for broader or custom use.

In multidisciplinary fields, standardization can often be difficult to implement, therefore, the reporting standard is also accompanied with an associated platform-specific XML schema. Although XML is not inherently user friendly, and is highly computationally amendable, the schema was specifically designed with the REDCap platform in mind. It is therefore immediately implementable to promote user friendliness in terms of both data capturing and governance, allowing accurate and seamless duplication in the local setting (Eito-Brun, 2018). Additionally, the accompanying Recommendations For Use guideline (in supplementary material) further enables use and user friendliness. The XML has been used extensively for describing data in many applications for storage or transport (Eito-Brun, 2018). The language, by its design, allows for extensibility and self-description. Its openly documented standards, wide adoption, and support in many applications and existing tools make it a good first choice for describing scientific data that could be exchanged between healthcare systems (Eito-Brun, 2018). It has previously been used in health reporting for such purposes (Huser et al., 2015; Schweiger et al., 2005).

As previously exhibited in oncology research, widespread utilization of the developed reporting guideline can function to reduce data and reporting inconsistency and redundancy across systems, as well as promote collaboration and(or) interoperability between systems (Biology et al., 2000; Hartwell et al., 2012; MacCarthy et al., 2018). Promoting such broad use could allow for improved data mapping in clinical registries, improving data quality and interoperability (Rastegar-Mojarad et al., 2017). A given standard may be more widely adopted if advocated or endorsed by "omics" databases, funding bodies and scientific journals, geared towards stroke research, specifically. To promote the adoption of the reporting guideline, we hope to employ the reporting guideline within our own consortia studies, and advocate use on an international platform.

In the future, the H3ABioNet's Data & Standards Work Package aims to develop more domain-specific reporting guidelines which are relevant to both African health and the H3Africa consortia. We also aim to align our efforts with the standardization efforts driven by GA4GH. This will include further refining elements such as ethnicity, diet and prescribed medicine to accommodate African-specific considerations. The Minimum Information Required Guideline: Stroke Research and Clinical Data Reporting aims to promote FAIR reporting and will therefore be added to the FAIRsharing database, as the database provides curation support to resource maintainers, as well as a point of contact for the standard, and related support material (Wilkinson et al., 2016). Bearing in mind the diverse target group the reporting standard aims to accommodate, various methods of implementation will be investigated in the future, to provide comprehensive solutions for collaborative efforts and increase the research data value. Education and training in the use and implementation of these standards will be of high importance to supplement use. Furthermore, additional elements will be investigated for incorporation into the standard, including various environmental factors. Ultimately, the reporting guideline has the potential to support both the H3Africa community as well as the stroke research community at large with current and future research.

Data Accessibility Statement

All data referenced in the article can be found in Supplementary File 2.

Additional Files

The additional files for this article can be found as follows:

- **File 1.** Stroke Online Survey. DOI: https://doi.org/10.5334/dsj-2019-026.s1
- **File 2.** Stroke Online Survey – Raw Results. DOI: https://doi.org/10.5334/dsj-2019-026.s2
- **File 3.** The Minimum Information Required Guideline: Stroke Research and Clinical Data Reporting Data Dictionary (Version 1.0). DOI: https://doi.org/10.5334/dsj-2019-026.s3
- **File 4.** Recommendations For Use Guideline. DOI: https://doi.org/10.5334/dsj-2019-026.s4
- **File 5.** Stroke Research Data Reporting REDCap XML. DOI: https://doi.org/10.5334/dsj-2019-026.s5
- **File 6.** Relationship between the different sections of the reporting guideline. DOI: https://doi.org/10.5334/dsj-2019-026.s6

Acknowledgements

The authors would like to acknowledge the stroke researchers, clinicians and hospital departments who participated and contributed to the consolidation and finalization of the reporting guideline.

Funding Information

H3ABioNet is supported by the National Institutes of Health Common Fund under grant number U41HG006941. Stroke Investigative Research & Educational Network (SIREN) U54HG007479 and Systematic Investigation of Blacks with Stroke using Genomics (SIBS Genomics) R01NS107900 SIBS Genomics are supported by grants from the National Institutes of Health. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

The current project was led by Judit Kumuthini and Mayowa Owolabi and supervised by Nicola Mulder. Lyndon Zass and Melek Chaouch were the primary developers of the reporting guidelines, along with significant contributions from Paul Olowoyo, Faniyan Moyinoluwalogo, Gordon Wells and Victoria Nembaware. Michael Thompson and Mamana Mbiyavanga were the primary developers of the XML files.

References

Adeloye, D. 2014. An estimate of the incidence and prevalence of stroke in Africa: a systematic review and meta-analysis. *PloS One*, 9: e100724. DOI: https://doi.org/10.1371/journal.pone.0100724

Akinyemi, RO, Ovbiagele, B, Akpalu, A, Jenkins, C, Sagoe, K, Owolabi, L, Sarfo, F, Obiako, R, Gebreziabher, M, Melikam, E, Warth, S, Arulogun, O, Lackland, D, Ogunniyi, A, Tiwari, H, Kalaria, RN, Arnett, D, Owolabi, MO and **SIREN Investigators as Members of the H3Africa Consortium.** 2015. Stroke genomics in people of African ancestry: charting new paths. *Cardiovasc. J. Afr*, 26: S39–49. DOI: https://doi.org/10.5830/CVJA-2015-039

Benjamin, EJ, Blaha, MJ, Chiuve, SE, Cushman, M, Das, SR, Deo, R, de Ferranti, SD, Floyd, J, Fornage, M, Gillespie, C, Isasi, CR, Jiménez, MC, Jordan, LC, Judd, SE, Lackland, D, Lichtman, JH, Lisabeth, L, Liu, S, Longenecker, CT, Mackey, RH, Matsushita, K, Mozaffarian, D, Mussolino, ME, Nasir, K, Neumar, RW, Palaniappan, L, Pandey, DK, Thiagarajan, RR, Reeves, MJ, Ritchey, M, Rodriguez, CJ, Roth, GA, Rosamond, WD, Sasson, C, Towfighi, A, Tsao, CW, Turner, MB, Virani, SS, Voeks, JH, Willey, JZ, Wilkins, JT, Wu, JH, Alger, HM, Wong, SS, Muntner, P and **American Heart Association Statistics Committee and Stroke Statistics Subcommittee.** 2017. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation*, 135: e146–e603. DOI: https://doi.org/10.1161/CIR.0000000000000485

Biology, NRC, (US) B. on, Pool, R, Esnayra, J. 2000. Barriers to the Use of Databases. National Academies Press (US).

Côté, RG, Jones, P, Apweiler, R and **Hermjakob, H.** 2006. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, 7: 97. DOI: https://doi.org/10.1186/1471-2105-7-97

Eito-Brun, R. 2018. Chapter 1 – XML: The Basis of the Language. In: Eito-Brun, R. (Ed.), *XML-Based Content Management*. Chandos Publishing, pp. 1–30. DOI: https://doi.org/10.1016/B978-0-08-100204-9.00001-9

H3Africa Consortium, Rotimi, C, Abayomi, A, Abimiku, A, Adabayeri, VM, Adebamowo, C, Adebiyi, E, Ademola, AD, Adeyemo, A, Adu, D, Affolabi, D, Agongo, G, Ajayi, S, Akarolo-Anthony, S, Akinyemi, R, Akpalu, A, Alberts, M, Alonso Betancourt, O, Alzohairy, AM, Ameni, G, Amodu, O, Anabwani, G, Andersen, K, Arogundade, F, Arulogun, O, Asogun, D, Bakare, R, Balde, N, Baniecki, ML, Beiswanger, C, Benkahla, A, Bethke, L, Boehnke, M, Boima, V, Brandful, J, Brooks, AI, Brosius, FC, Brown, C, Bucheton, B, Burke, DT, Burnett, BG, Carrington-Lawrence, S, Carstens, N, Chisi, J, Christoffels, A, Cooper, R, Cordell, H, Crowther, N, Croxton, T, de Vries, J, Derr, L, Donkor, P, Doumbia, S, Duncanson, A, Ekem, I, El Sayed, A, Engel, ME, Enyaru, JCK, Everett, D, Fadlelmola, FM, Fakunle, E, Fischbeck, KH, Fischer, A, Folarin, O, Gamieldien, J, Garry, RF, Gaseitsiwe, S, Gbadegesin, R, Ghansah, A, Giovanni, M, Goesbeck, P, Gomez-Olive, FX, Grant, DS, Grewal, R, Guyer, M, Hanchard, NA, Happi, CT, Hazelhurst, S, Hennig, BJ, Hertz-, C, Fowler, null, Hide, W, Hilderbrandt, F, Hugo-Hamman, C, Ibrahim, ME, James, R, Jaufeerally-Fakim, Y, Jenkins, C, Jentsch, U, Jiang, P-P, Joloba, M, Jongeneel, V, Joubert, F, Kader, M, Kahn, K, Kaleebu, P, Kapiga, SH, Kassim, SK, Kasvosve, I, Kayondo, J, Keavney, B, Kekitiinwa, A, Khan, SH, Kimmel, P, King, M-C, Kleta, R, Koffi, M, Kopp, J, Kretzler, M, Kumuthini, J, Kyobe, S, Kyobutungi, C, Lackland, DT, Lacourciere, KA, Landouré, G, Lawlor, R, Lehner, T, Lesosky, M, Levitt, N, Littler, K, Lombard, Z, Loring, JF, Lyantagaye, S, Macleod, A, Madden, EB, Mahomva, CR, Makani, J, Mamven, M, Marape, M, Mardon, G, Marshall, P, Martin, DP, Masiga, D, Mason, R, Mate-Kole, M, Matovu, E, Mayige, M, Mayosi, BM, Mbanya, JC, McCurdy, SA, McCarthy, MI, McIlleron, H, Mc'Ligeyo, SO, Merle, C, Mocumbi, AO, Mondo, C, Moran, JV, Motala, A, Moxey-Mims, M, Mpoloka, WS, Msefula, CL, Mthiyane, T, Mulder, N, Mulugeta, G, her, Mumba, D, Musuku, J, Nagdee, M, Nash, O, Ndiaye, D, Nguyen, AQ, Nicol, M, Nkomazana, O, Norris, S, Nsangi, B, Nyarko, A, Nyirenda, M, Obe, E, Obiakor, R, Oduro, A, Ofori-Acquah, SF, Ogah, O, Ogendo, S, Ohene-Frempong, K, Ojo, A, Olanrewaju, T, Oli, J, Osafo, C, Ouwe Missi Oukem-Boyer, O, Ovbiagele, B, Owen, A, Owolabi, MO, Owolabi, L, Owusu-Dabo, E, Pare, G, Parekh, R, Patterton, HG, Penno, MB, Peterson, J, Pieper, R, Plange-Rhule, J, Pollak, M, Puzak, J, Ramesar, RS, Ramsay, M, Rasooly, R, Reddy, S, Sabeti, PC, Sagoe, K, Salako, T, Samassékou, O, Sandhu, MS, Sankoh, O, Sarfo, FS, Sarr, M, Shaboodien, G, Sidibe, I, Simo, G, Simuunza, M, Smeeth, L, Sobngwi, E, Soodyall, H, Sorgho, H, Sow Bah, O, Srinivasan, S, Stein, DJ, Susser, ES, Swanepoel, C, Tangwa, G, Tareila, A, Tastan Bishop, O, Tayo, B, Tiffin, N, Tinto, H, Tobin, E, Tollman, SM, Traoré, M, Treadwell, MJ, Troyer, J, Tsimako-Johnstone, M, Tukei, V, Ulasi, I, Ulenga, N, van Rooyen, B, Wachinou, AP, Waddy, SP, Wade, A, Wayengera, M, Whitworth, J, Wideroff, L, Winkler, CA, Winnicki, S, Wonkam, A, Yewondwos, M, Sen, T, Yozwiak, N and **Zar, H.** 2014. Research capacity. Enabling the genomic revolution in Africa. *Science*, 344: 1346–1348. DOI: https://doi.org/10.1126/science.1251546

Hamilton, CM, Strader, LC, Pratt, JG, Maiese, D, Hendershot, T, Kwok, RK, Hammond, JA, Huggins, W, Jackman, D, Pan, H, Nettles, DS, Beaty, TH, Farrer, LA, Kraft, P, Marazita, ML, Ordovas, JM, Pato, CN, Spitz, MR, Wagener, D, Williams, M, Junkins, HA, Harlan, WR, Ramos, EM and **Haines, J.** 2011. The PhenX Toolkit: Get the Most From Your Measures. *Am. J. Epidemiol*, 174: 253–260. DOI: https://doi.org/10.1093/aje/kwr193

Harris, PA, Taylor, R, Thielke, R, Payne, J, Gonzalez, N and **Conde, JG.** 2009. Research Electronic Data Capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform*, 42: 377–381. DOI: https://doi.org/10.1016/j.jbi.2008.08.010

Hartwell, SW, Deng, X, Fisher, W, Fulwiler, C, Sambamoorthi, U, Johnson, C, Pinals, DA, Sampson, L and **Siegfriedt, J.** 2012. Harmonizing databases? Developing a quasi-experimental design to evaluate a public mental health re-entry program. *Eval. Program Plann., Collaborative Evaluation: Theory and Practice*, 35: 461–472. DOI: https://doi.org/10.1016/j.evalprogplan.2012.02.001

Huser, V, Sastry, C, Breymaier, M, Idriss, A and **Cimino, JJ.** 2015. Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). *J. Biomed. Inform.* 57: 88–99. DOI: https://doi.org/10.1016/j.jbi.2015.06.023

Khan, AM. 2017. Guidelines for standardizing and increasing the transparency in the reporting of biomedical research. *J. Thorac. Dis*, 9: 2697–2702. DOI: https://doi.org/10.21037/jtd.2017.07.30

Kumuthini, J, Mbiyavanga, M, Chimusa, ER, Pathak, J, Somervuo, P, Van Schaik, RH, Dolzan, V, Mizzi, C, Kalideen, K, Ramesar, RS, Macek, M, Patrinos, GP and **Squassina, A.** 2016. Minimum information required for a DMET experiment reporting. *Pharmacogenomics*, 17: 1533–1545. DOI: https://doi.org/10.2217/pgs-2016-0015

Lee, CH and **Yoon, H-J.** 2017. Medical big data: promise and challenges. *Kidney Res. Clin. Pract*, 36: 3–11. DOI: https://doi.org/10.23876/j.krcp.2017.36.1.3

Luo, J, Wu, M, Gopukumar, D and **Zhao, Y.** 2016. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed. Inform. Insights*, 8: 1–10. DOI: https://doi.org/10.4137/BII.S31559

MacCarthy, A, Kirtley, S, de Beyer, JA, Altman, DG and **Simera, I.** 2018. Reporting guidelines for oncology research: helping to maximise the impact of your research. *Br. J. Cancer*, 118: 619–628. DOI: https://doi.org/10.1038/bjc.2017.407

Mulder, NJ, Adebiyi, E, Alami, R, Benkahla, A, Brandful, J, Doumbia, S, Everett, D, Fadlelmola, FM, Gaboun, F, Gaseitsiwe, S, Ghazal, H, Hazelhurst, S, Hide, W, Ibrahimi, A, Jaufeerally Fakim, Y, Jongeneel, CV, Joubert, F, Kassim, S, Kayondo, J, Kumuthini, J, Lyantagaye, S, Makani, J, Mansour Alzohairy, A, Masiga, D, Moussa, A, Nash, O, Ouwe Missi Oukem-Boyer, O, Owusu-Dabo, E, Panji, S, Patterton, H, Radouani, F, Sadki, K, Seghrouchni, F, Tastan Bishop, Ö, Tiffin, N and **Ulenga, N.** 2016. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Res*, 26: 271–277. DOI: https://doi.org/10.1101/gr.196295.115

Noy, NF, Shah, NH, Whetzel, PL, Dai, B, Dorf, M, Griffith, N, Jonquet, C, Rubin, DL, Storey, M-A, Chute, CG and **Musen, MA.** 2009. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res*, 37: W170–W173. DOI: https://doi.org/10.1093/nar/gkp440

Owolabi, MO, Sarfo, F, Akinyemi, R, Gebregziabher, M, Akpa, O, Akpalu, A, Wahab, K, Obiako, R, Owolabi, L and **Ovbiagele, B.** 2018. Dominant modifiable risk factors for stroke in Ghana and Nigeria (SIREN): a case-control study. *Lancet Glob. Health*, 6: e436–e446. DOI: https://doi.org/10.1016/S2214-109X(18)30002-0

Rastegar-Mojarad, M, Sohn, S, Wang, L, Shen, F, Bleeker, TC, Cliby, WA and **Liu, H.** 2017. Need of informatics in designing interoperable clinical registries. *Int. J. Med. Inf*, 108: 78–84. DOI: https://doi.org/10.1016/j.ijmedinf.2017.10.004

Robinson, PN and **Mundlos, S.** 2010. The human phenotype ontology. *Clin. Genet*, 77: 525–534. DOI: https://doi.org/10.1111/j.1399-0004.2010.01436.x

Sacco, RL, Kasner, SE, Broderick, JP, Caplan, LR, Connors, JJB, Culebras, A, Elkind, MSV, George, MG, Hamdan, AD, Higashida, RT, Hoh, BL, Janis, LS, Kase, CS, Kleindorfer, DO, Lee, J-M, Moseley, ME, Peterson, ED, Turan, TN, Valderrama, AL, Vinters, HV, American Heart Association Stroke Council, Council on Cardiovascular Surgery and Anesthesia, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular and Stroke Nursing, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease and **Council on Nutrition, Physical Activity and Metabolism.** 2013. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 44: 2064–2089. DOI: https://doi.org/10.1161/STR.0b013e318296aeca

Schweiger, R, Brumhard, M, Hoelzer, S and **Dudeck, J.** 2005. Implementing health care systems using XML standards. *Int. J. Med. Inf*, 74: 267–277. DOI: https://doi.org/10.1016/j.ijmedinf.2004.04.019

Skrocki, M. 2013. Standardization Needs for Effective Interoperability. *Trans. Int. Conf. Health Inf. Technol. Adv*.

Taylor, CF, Paton, NW, Lilley, KS, Binz, P-A, Julian, RK, Jones, AR, Zhu, W, Apweiler, R, Aebersold, R, Deutsch, EW, Dunn, MJ, Heck, AJR, Leitner, A, Macht, M, Mann, M, Martens, L, Neubert, TA, Patterson, SD, Ping, P, Seymour, SL, Souda, P, Tsugita, A, Vandekerckhove, J, Vondriska, TM, Whitelegge, JP, Wilkins, MR, Xenarios, I, Yates, JR and **Hermjakob, H.** 2007. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 25: 887–893. DOI: https://doi.org/10.1038/nbt1329

Wang, H, Naghavi, M, Allen, C, Barber, RM, Bhutta, ZA, Carter, A, Casey, DC, Charlson, FJ, Chen, AZ, Coates, MM, Coggeshall, M, Dandona, L, Dicker, DJ, Erskine, HE, Ferrari, AJ, Fitzmaurice, C, Foreman, K, Forouzanfar, MH, Fraser, MS, Fullman, N, Gething, PW, Goldberg, EM, Graetz, N, Haagsma, JA, Hay, SI, Huynh, C, Johnson, CO, Kassebaum, NJ, Kinfu, Y, Kulikoff, XR, Kutz, M, Kyu, HH, Larson, HJ, Leung, J, Liang, X, Lim, SS, Lind, M, Lozano, R, Marquez, N, Mensah, GA, Mikesell, J, Mokdad, AH, Mooney, MD, Nguyen, G, Nsoesie, E, Pigott, DM, Pinho, C, Roth, GA, Salomon, JA, Sandar, L, Silpakit, N, Sligar, A, Sorensen, RJD, Stanaway, J, Steiner, C, Teeple, S, Thomas, BA,

Troeger, C, VanderZanden, A, Vollset, SE, Wanga, V, Whiteford, HA, Wolock, T, Zoeckler, L, Abate, KH, Abbafati, C, Abbas, KM, Abd-Allah, F, Abera, SF, Abreu, DMX, Abu-Raddad, LJ, Abyu, GY, Achoki, T, Adelekan, AL, Ademi, Z, Adou, AK, Adsuar, JC, Afanvi, KA, Afshin, A, Agardh, EE, Agarwal, A, Agrawal, A, Kiadaliri, AA, Ajala, ON, Akanda, AS, Akinyemi, RO, Akinyemiju, TF, Akseer, N, Lami, FHA, Alabed, S, Al-Aly, Z, Alam, K, Alam, NKM, Alasfoor, D, Aldhahri, SF, Aldridge, RW, Alegretti, MA, Aleman, AV, Alemu, ZA, Alexander, LT, Alhabib, S, Ali, R, Alkerwi, A, Alla, F, Allebeck, P, Al-Raddadi, R, Alsharif, U, Altirkawi, KA, Martin, EA, Alvis-Guzman, N, Amare, AT, Amegah, AK, Ameh, EA, Amini, H, Ammar, W, Amrock, SM, Andersen, HH, Anderson, BO, Anderson, GM, Antonio, CAT, Aregay, AF, Ärnlöv, J, Arsenijevic, VSA, Artaman, A, Asayesh, H, Asghar, RJ, Atique, S, Avokpaho, EFGA, Awasthi, A, Azzopardi, P, Bacha, U, Badawi, A, Bahit, MC, Balakrishnan, K, Banerjee, A, Barac, A, Barker-Collo, SL, Bärnighausen, T, Barregard, L, Barrero, LH, Basu, A, Basu, S, Bayou, YT, Bazargan-Hejazi, S, Beardsley, J, Bedi, N, Beghi, E, Belay, HA, Bell, B, Bell, ML, Bello, AK, Bennett, DA, Bensenor, IM, Berhane, A, Bernabé, E, Betsu, BD, Beyene, AS, Bhala, N, Bhalla, A, Biadgilign, S, Bikbov, B, Abdulhak, AAB, Biroscak, BJ, Biryukov, S, Bjertness, E, Blore, JD, Blosser, CD, Bohensky, MA, Borschmann, R, Bose, D, Bourne, RRA, Brainin, M, Brayne, CEG, Brazinova, A, Breitborde, NJK, Brenner, H, Brewer, JD, Brown, A, Brown, J, Brugha, TS, Buckle, GC, Butt, ZA, Calabria, B, Campos-Nonato, IR, Campuzano, JC, Carapetis, JR, Cárdenas, R, Carpenter, DO, Carrero, JJ, Castañeda-Orjuela, CA, Rivas, JC, Catalá-López, F, Cavalleri, F, Cercy, K, Cerda, J, Chen, W, Chew, A, Chiang, PP-C, Chibalabala, M, Chibueze, CE, Chimed-Ochir, O, Chisumpa, VH, Choi, J-YJ, Chowdhury, R, Christensen, H, Christopher, DJ, Ciobanu, LG, Cirillo, M, Cohen, AJ, Colistro, V, Colomar, M, Colquhoun, SM, Cooper, C, Cooper, LT, Cortinovis, M, Cowie, BC, Crump, JA, Damsere-Derry, J, Danawi, H, Dandona, R, Daoud, F, Darby, SC, Dargan, PI, das Neves, J, Davey, G, Davis, AC, Davitoiu, DV, de Castro, EF, de Jager, P, Leo, DD, Degenhardt, L, Dellavalle, RP, Deribe, K, Deribew, A, Dharmaratne, SD, Dhillon, PK, Diaz-Torné, C, Ding, EL, Santos, KPB, dos Dossou, E, Driscoll, TR, Duan, L, Dubey, M, Duncan, BB, Ellenbogen, RG, Ellingsen, CL, Elyazar, I, Endries, AY, Ermakov, SP, Eshrati, B, Esteghamati, A, Estep, K, Faghmous, IDA, Fahimi, S, Faraon, EJA, Farid, TA, Farinha, CSeS, Faro, A, Farvid, MS, Farzadfar, F, Feigin, VL, Fereshtehnejad, S-M, Fernandes, JG, Fernandes, JC, Fischer, F, Fitchett, JRA, Flaxman, A, Foigt, N, Fowkes, FGR, Franca, EB, Franklin, RC, Friedman, J, Frostad, J, Fürst, T, Futran, ND, Gall, SL, Gambashidze, K, Gamkrelidze, A, Ganguly, P, Gankpé, FG, Gebre, T, Gebrehiwot, TT, Gebremedhin, AT, Gebru, AA, Geleijnse, JM, Gessner, BD, Ghoshal, AG, Gibney, KB, Gillum, RF, Gilmour, S, Giref, AZ, Giroud, M, Gishu, MD, Giussani, G, Glaser, E, Godwin, WW, Gomez-Dantes, H, Gona, P, Goodridge, A, Gopalani, SV, Gosselin, RA, Gotay, CC, Goto, A, Gouda, HN, Greaves, F, Gugnani, HC, Gupta, R, Gupta, R, Gupta, V, Gutiérrez, RA, Hafezi-Nejad, N, Haile, D, Hailu, AD, Hailu, GB, Halasa, YA, Hamadeh, RR, Hamidi, S, Hancock, J, Handal, AJ, Hankey, GJ, Hao, Y, Harb, HL, Harikrishnan, S, Haro, JM, Havmoeller, R, Heckbert, SR, Heredia-Pi, IB, Heydarpour, P, Hilderink, HBM, Hoek, HW, Hogg, RS, Horino, M, Horita, N, Hosgood, HD, Hotez, PJ, Hoy, DG, Hsairi, M, Htet, AS, Htike, MMT, Hu, G, Huang, C, Huang, H, Huiart, L, Husseini, A, Huybrechts, I, Huynh, G, Iburg, KM, Innos, K, Inoue, M, Iyer, VJ, Jacobs, TA, Jacobsen, KH, Jahanmehr, N, Jakovljevic, MB, James, P, Javanbakht, M, Jayaraman, SP, Jayatilleke, AU, Jeemon, P, Jensen, PN, Jha, V, Jiang, G, Jiang, Y, Jibat, T, Jimenez-Corona, A, Jonas, JB, Joshi, TK, Kabir, Z, Kamal, R, Kan, H, Kant, S, Karch, A, Karema, CK, Karimkhani, C, Karletsos, D, Karthikeyan, G, Kasaeian, A, Katibeh, M, Kaul, A, Kawakami, N, Kayibanda, JF, Keiyoro, PN, Kemmer, L, Kemp, AH, Kengne, AP, Keren, A, Kereselidze, M, Kesavachandran, CN, Khader, YS, Khalil, IA, Khan, AR, Khan, EA, Khang, Y-H, Khera, S, Khoja, TAM, Kieling, C, Kim, D, Kim, YJ, Kissela, BM, Kissoon, N, Knibbs, LD, Knudsen, AK, Kokubo, Y, Kolte, D, Kopec, JA, Kosen, S, Koul, PA, Koyanagi, A, Krog, NH, Defo, BK, Bicer, BK, Kudom, AA, Kuipers, EJ, Kulkarni, VS, Kumar, GA, Kwan, GF, Lal, A, Lal, DK, Lalloo, R, Lallukka, T, Lam, H, Lam, JO, Langan, SM, Lansingh, VC, Larsson, A, Laryea, DO, Latif, AA, Lawrynowicz, AEB, Leigh, J, Levi, M, Li, Y, Lindsay, MP, Lipshultz, SE, Liu, PY, Liu, S, Liu, Y, Lo, L-T, Logroscino, G, Lotufo, PA, Lucas, RM, Lunevicius, R, Lyons, RA, Ma, S, Machado, VMP, Mackay, MT, MacLachlan, JH, Razek, HMAE, Magdy, M, Razek, AE, Majdan, M, Majeed, A, Malekzadeh, R, Manamo, WAA, Mandisarisa, J, Mangalam, S, Mapoma, CC, Marcenes, W, Margolis, DJ, Martin, GR, Martinez-Raga, J, Marzan, MB, Masiye, F, Mason-Jones, AJ, Massano, J, Matzopoulos, R, Mayosi, BM, McGarvey, ST, McGrath, JJ, McKee, M, McMahon, BJ, Meaney, PA, Mehari, A, Mehndiratta, MM, Mejia-Rodriguez, F, Mekonnen, AB, Melaku, YA, Memiah, P, Memish, ZA, Mendoza, W, Meretoja, A, Meretoja, TJ, Mhimbira, FA, Micha, R, Millear, A, Miller, TR, Mirarefin, M, Misganaw, A, Mock, CN, Mohammad,

KA, Mohammadi, A, Mohammed, S, Mohan, V, Mola, GLD, Monasta, L, Hernandez, JCM, Montero, P, Montico, M, Montine, TJ, Moradi-Lakeh, M, Morawska, L, Morgan, K, Mori, R, Mozaffarian, D, Mueller, UO, Murthy, GVS, Murthy, S, Musa, KI, Nachega, JB, Nagel, G, Naidoo, KS, Naik, N, Naldi, L, Nangia, V, Nash, D, Nejjari, C, Neupane, S, Newton, CR, Newton, JN, Ng, M, Ngalesoni, FN, Ngirabega, JdeD, Nguyen, QL, Nisar, MI, Pete, PMN, Nomura, M, Norheim, OF, Norman, PE, Norrving, B, Nyakarahuka, L, Ogbo, FA, Ohkubo, T, Ojelabi, FA, Olivares, PR, Olusanya, BO, Olusanya, JO, Opio, JN, Oren, E, Ortiz, A, Osman, M, Ota, E, Ozdemir, R, Pa, M, Pain, A, Pandian, JD, Pant, PR, Papachristou, C, Park, E-K, Park, J-H, Parry, CD, Parsaeian, M, Caicedo, AJP, Patten, SB, Patton, GC, Paul, VK, Pearce, N, Pedro, JM, Stokic, LP, Pereira, DM, Perico, N, Pesudovs, K, Petzold, M, Phillips, MR, Piel, FB, Pillay, JD, Plass, D, Platts-Mills, JA, Polinder, S, Pope, CA, Popova, S, Poulton, RG, Pourmalek, F, Prabhakaran, D, Qorbani, M, Quame-Amaglo, J, Quistberg, DA, Rafay, A, Rahimi, K, Rahimi-Movaghar, V, Rahman, M, Rahman, MHU, Rahman, SU, Rai, RK, Rajavi, Z, Rajsic, S, Raju, M, Rakovac, I, Rana, SM, Ranabhat, CL, Rangaswamy, T, Rao, P, Rao, SR, Refaat, AH, Rehm, J, Reitsma, MB, Remuzzi, G, Resnikoff, S, Ribeiro, AL, Ricci, S, Blancas, MJR, Roberts, B, Roca, A, Rojas-Rueda, D, Ronfani, L, Roshandel, G, Rothenbacher, D, Roy, A, Roy, NK, Ruhago, GM, Sagar, R, Saha, S, Sahathevan, R, Saleh, MM, Sanabria, JR, Sanchez-Niño, MD, Sanchez-Riera, L, Santos, IS, Sarmiento-Suarez, R, Sartorius, B, Satpathy, M, Savic, M, Sawhney, M, Schaub, MP, Schmidt, MI, Schneider, IJC, Schöttker, B, Schutte, AE, Schwebel, DC, Seedat, S, Sepanlou, SG, Servan-Mori, EE, Shackelford, KA, Shaddick, G, Shaheen, A, Shahraz, S, Shaikh, MA, Shakh-Nazarova, M, Sharma, R, She, J, Sheikhbahaei, S, Shen, J, Shen, Z, Shepard, DS, Sheth, KN, Shetty, BP, Shi, P, Shibuya, K, Shin, M-J, Shiri, R, Shiue, I, Shrime, MG, Sigfusdottir, ID, Silberberg, DH, Silva, DAS, Silveira, DGA, Silverberg, JI, Simard, EP, Singh, A, Singh, GM, Singh, JA, Singh, OP, Singh, PK, Singh, V, Soneji, S, Søreide, K, Soriano, JB, Sposato, LA, Sreeramareddy, CT, Stathopoulou, V, Stein, DJ, Stein, MB, Stranges, S, Stroumpoulis, K, Sunguya, BF, Sur, P, Swaminathan, S, Sykes, BL, Szoeke, CEI, Tabarés-Seisdedos, R, Tabb, KM, Takahashi, K, Takala, JS, Talongwa, RT, Tandon, N, Tavakkoli, M, Taye, B, Taylor, HR, Ao, BJT, Tedla, BA, Tefera, WM, Have, MT, Terkawi, AS, Tesfay, FH, Tessema, GA, Thomson, AJ, Thorne-Lyman, AL, Thrift, AG, Thurston, GD, Tillmann, T, Tirschwell, DL, Tonelli, M, Topor-Madry, R, Topouzis, F, Towbin, JA, Traebert, J, Tran, BX, Truelsen, T, Trujillo, U, Tura, AK, Tuzcu, EM, Uchendu, US, Ukwaja, KN, Undurraga, EA, Uthman, OA, Dingenen, RV, van Donkelaar, A, Vasankari, T, Vasconcelos, AMN, Venketasubramanian, N, Vidavalur, R, Vijayakumar, L, Villalpando, S, Violante, FS, Vlassov, VV, Wagner, JA, Wagner, GR, Wallin, MT, Wang, L, Watkins, DA, Weichenthal, S, Weiderpass, E, Weintraub, RG, Werdecker, A, Westerman, R, White, RA, Wijeratne, T, Wilkinson, JD, Williams, HC, Wiysonge, CS, Woldeyohannes, SM, Wolfe, CDA, Won, S, Wong, JQ, Woolf, AD, Xavier, D, Xiao, Q, Xu, G, Yakob, B, Yalew, AZ, Yan, LL, Yano, Y, Yaseri, M, Ye, P, Yebyo, HG, Yip, P, Yirsaw, BD, Yonemoto, N, Yonga, G, Younis, MZ, Yu, S, Zaidi, Z, Zaki, MES, Zannad, F, Zavala, DE, Zeeb, H, Zeleke, BM, Zhang, H, Zodpey, S, Zonies, D, Zuhlke, LJ, Vos, T, Lopez, AD and **Murray, CJL.** 2016. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388: 1459–1544. DOI: https://doi.org/10.1016/S0140-6736(16)31012-1

Wang, L and **Alexander, CA.** 2016. Stroke Care and the Role of Big Data in Healthcare and Stroke. *Rehabil. Sci*, 1: 16. DOI: https://doi.org/10.11648/j.rs.20160101.13

Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, da Silva Santos, LB, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJG, Groth, P, Goble, C, Grethe, JS, Heringa, J, Hoen, PAC 't, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Rocca-Serra, P, Roos, M, van Schaik, R, Sansone, S-A, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, MA, Thompson, M, van der Lei, J, van Mulligen, E, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J and **Mons, B.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3: sdata201618. DOI: https://doi.org/10.1038/sdata.2016.18

How to cite this article: Kumuthini, J, Zass, L, Chaouch, M, Thompson, M, Olowoyo, P, Mbiyavanga, M, Moyinoluwalogo, F, Wells, G, Nembaware, V, Mulder, NJ, Owolabi, M and H3ABioNet Consortium's Data and Standards Working Group as members of the H3Africa Consortium. 2019. Proposed Guideline for Minimum Information Stroke Research and Clinical Data Reporting. *Data Science Journal,* 18: 26, pp. 1–18. DOI: https://doi.org/10.5334/dsj-2019-026

Submitted: 30 January 2019 **Accepted:** 07 May 2019 **Published:** 27 June 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS