BMJ Open

Proposed minimum information guideline for kidney disease—research and clinical data reporting: a cross-sectional study

Judit Kumuthini,[1] Christiaan van Woerden,[2,3] Andrew Mallett,[4,5,6] Lyndon Zass,[1] Melek Chaouch,[7] Michael Thompson,[8] Katherine Johnston,[2] Mamana Mbiyavanga,[2] Shakuntala Baichoo,[9] Zahra Mungloo-Dilmohamud,[9] Chirag Patel,[4,6] Nicola Mulder[10]

To cite: Kumuthini J, van Woerden C, Mallett A, *et al*. Proposed minimum information guideline for kidney disease—research and clinical data reporting: a cross-sectional study. *BMJ Open* 2019;**9**:e029539. doi:10.1136/bmjopen-2019-029539

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-029539).

Received 31 January 2019
Revised 07 October 2019
Accepted 11 October 2019

For numbered affiliations see end of article.

Correspondence to
Dr Judit Kumuthini;
jkumuthini@gmail.com

ABSTRACT

Objective This project aimed to develop and propose a standardised reporting guideline for kidney disease research and clinical data reporting, in order to improve kidney disease data quality and integrity, and combat challenges associated with the management and challenges of 'Big Data'.

Methods A list of recommendations was proposed for the reporting guideline based on the systematic review and consolidation of previously published data collection and reporting standards, including PhenX measures and Minimal Information about a Proteomics Experiment (MIAPE). Thereafter, these recommendations were reviewed by domain-specialists using an online survey, developed in Research Electronic Data Capture (REDCap). Following interpretation and consolidation of the survey results, the recommendations were mapped to existing ontologies using Zooma, Ontology Lookup Service and the Bioportal search engine. Additionally, an associated eXtensible Markup Language schema was created for the REDCap implementation to increase user friendliness and adoption.

Results The online survey was completed by 53 respondents; the majority of respondents were dual clinician-researchers (57%), based in Australia (35%), Africa (33%) and North America (22%). Data elements within the reporting standard were identified as participant-level, study-level and experiment-level information, further subdivided into essential or optional information.

Conclusion The reporting guideline is readily employable for kidney disease research projects, and also adaptable for clinical utility. The adoption of the reporting guideline in kidney disease research can increase data quality and the value for long-term preservation, ensuring researchers gain the maximum benefit from their collected and generated data.

Strengths and limitations of this study

- The reporting guideline references and is mapped to standardised collection measures and biomedical ontologies found in the Phenotypes and eXposures Toolkit, the Ontology Lookup Service and Bioportal, respectively.
- The reporting guideline was reviewed by global domain-specialists.
- Limited survey feedback was received from Asian and South American countries.
- An eXtensible Markup Language (XML) schema was developed to promote user-friendliness and maintain the relationship between the reporting guideline's sub-sections.
- The XML schema does not inherently incorporate mapped ontologies; therefore, strategies are being investigated to automate the process.

INTRODUCTION

'Big Data' and bioinformatics have become crucial components of modern biomedical research and healthcare.[1] In biomedical research, 'Big Data', commonly characterised by volume and variety, refers to data sets which are too large or complex to be analysed using traditional methods, often requiring the use of computational analyses to derive biological meaning.[2] Biomedical 'Big Data' has many potential fields of application, including personalised medicine, predictive modelling and clinical decision support, and disease and safety surveillance.[3] In the field of nephrology, the use of 'Big Data' has led to the improved understanding of the pathological processes and the underlying aetiologies associated with kidney disease,[4] as well as the increased identification of genetic kidney diseases[5 6] and novel diagnostic methods for these diseases.[7]

Challenges associated with the data lifecycle, including the collection, management, storage and analysis of data, hamper the use and potential benefit of 'Big Data'. These factors are compounded by additional biomedical research challenges, such as the inability to recruit sufficient sample sizes, as well as the lack of research capacity, funding and infrastructure, especially in low-income regions.[8 9] Additionally, the use

of ill-defined ontologies, data dictionaries and data management plans, contribute to data incompatibility and prevents researchers from reaping the maximum benefit from their collected and generated data.[10]

Standardising clinical and research data collection, reporting, management or storage can combat these challenges, supporting the effective integration of 'Big Data' and bioinformatics in biomedical research,[11 12] enhancing data compatibility, interoperability, reproducibility and reuse,[12] and facilitating data sharing and collaboration.[11] The use of biomedical reporting standards and ontologies facilitate data standardisation by promoting the use of or adherence to common terminology and(or) reporting criteria.[10 13] To this end, several initiatives have been driving data standardisation efforts in biomedical research. The consensus measures for Phenotypes and eXposures (PhenX) Toolkit (www.phenxtoolkit.org), has proposed phenotype collection tools for harmonised data collection, although some tools are limited in terms of applicability to low-resource settings.[14] Similar aims are being driven by FAIRsharing (www.fairsharing.org),[15] a dynamic standards database which aims to promote FAIR (findable, accessible, interoperable and reusable) principles,[16] and the Global Alliance for Genomics and Health (www.ga4gh.org), a policy-framing and technical standards-setting organisation. Multiple kidney-associated ontologies which define known kidney diseases and assist routine data studies and case identification have previously been developed, including the chronic kidney disease ontology,[17] as well as the renal subsections in the gene ontology, the human phenotype ontology[18] and the Systematized Nomenclature of Medicine-Clinical Terms.[19] However, no reporting guideline has previously been constructed for kidney disease clinical and research data reporting.

The Human Heredity and Health in Africa's (H3Africa) Bioinformatics Network's (H3ABioNet, www.h3abionet.org)[20 21] Data & Standards work package aims to develop domain-specific data reporting standards and data dictionaries, applicable to the H3Africa consortium, in order to specifically address the data management concerns in low-resource and low-income regions, affected by global health concerns, but lacking capacity to address these concerns. By consolidating and harmonising several published ontologies, collection standards and reporting standards, and consulting domain-specialists through an open online survey, the project drew from the experience of previous standardisation initiatives and aimed to develop a multipurpose reporting guideline which focused on the reporting of both clinical and research data within the kidney disease field, entitled, 'The Minimum Information Required Guideline for Kidney Disease: Research and Clinical Data Reporting (Version 1.0)'.

METHODS

Patient involvement

No patients or public were included in the methodology of the study. The survey employed in the study was strictly distributed to domain-specialists, hereafter defined.

Development of draft

Following the review of previously published literature and standards, several recommendations for the 'Minimum Information Required Guideline: Kidney Disease Research and Clinical Data Reporting' standards were proposed. The standards included were separated into two streams, the standards relevant to the collection of clinical data and those relevant to the reporting of research data. The standards relevant to clinical data collection included the H3Africa Standard Case Report Form (www.h3abionet.org/data-standards/datastds), the CKDO and various collection measures hosted on PhenX. The standards relevant to research data reporting included various experimental reporting guidelines hosted on FAIRsharing, such as MIAPE, MIDE, MIRAGE, MINSEQE, MIAME and more, from which common study-specific and experiment-specific elements were derived. Based on these recommendations, a reporting standard was drafted, which divided the proposed recommendations into three subsections; participant (patient), study-level and experiment-level information. The developed draft aimed to be both comprehensive and adaptable for both acquired and inherited kidney diseases, containing and querying elements specific to one or both types. Thereafter, recommendations (henceforth referred to as elements) were manually defined using ontologies found through the BioPortal search engine,[22] the Ontology Lookup Service at the European Bioinformatics Institute[23] and the Zooma annotation tool.

Online survey

To remove any existing reporting inconsistencies, domain-specialists, consisting of kidney disease researchers and clinicians, were consulted to review the proposed elements using an online survey. Domain-specialists were defined as both clinicians and researchers that have been involved in kidney disease research for at least a year, as part of an existing collaborative kidney disease research group or network (including the H3Africa Kidney Disease Research Network, the Australian KidGen Collaborative and Renal Genetics Flagships, Kidney Research UK and The Renal Network) and contacted via email. Domain-specialists were asked to evaluate, harmonise and consolidate the proposed elements, as well as identify which elements represented essential (E) or optional (O) information, and propose additional elements. Elements were classified as either E or O based on the E% percentage of E votes received. This percentage was calculated by dividing the total number of E votes by the number of votes made for a given element. Elements with lower than 50% were classified as O, while elements higher than 70% were classified as E, and elements within the 50–70 E% were classified with discretion based on correlations with the previously developed reporting guidelines and standard collection measures. Additional suggestions, not included in the draft, made by respondents were similarly classified.

Because the survey was constructed to be open, no limitations were set with regard to the number of participants,

and respondents were encouraged to distribute the survey within their own networks. The online survey was developed, and study data were collected and managed using Research Electronic Data Capture (REDCap),[24] hosted at The Centre for Proteomic and Genomic Research. The online survey consisted of 4 sections and 77 fields (online supplementary file 1). REDCap was employed for security and maintenance purposes.

Development of eXtensible Markup Language

To supplement usability and user-friendliness, an associated eXtensible Markup Language (XML) schema was designed to carry all the data and metadata within the reporting guideline and allow data exchange between dissimilar systems. The XML schema 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. Additionally, due to its user-friendliness and availability to research institutions worldwide, the XML schema was designed for implementation in REDCap.

RESULTS

The online survey was completed by 53 international domain specialists. Of these respondents, 29% were working as clinicians, 14% 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 (41%), while 37% of respondents had more than 20 years' experience in the field, 13% had more than 5 years' experience' in the field and 9% had less than 5 years' experience in the field. The majority of respondents were based in Australia (35%), followed by Africa (33%), North America (22%), Europe (9%) and Asia (2%). The raw survey results can be found in online supplementary file 2. Figures 1 and 2 illustrate the survey response to the proposed elements. Furthermore, respondents also proposed additional elements which shaped the final structure of the reporting guideline, including, but not limited to, congenital conditions, histopathology, language, physical activity and more.

The Minimum Information Required Guideline: Kidney Disease Research and Clinical Data Reporting is summarised in table 1.

The quintessential information reported using the standard can be separated into three fields; participant-level, study-level 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, kidney disease history, sample-specific information, kidney disease-related information, prescribed medication, non-prescribed medication

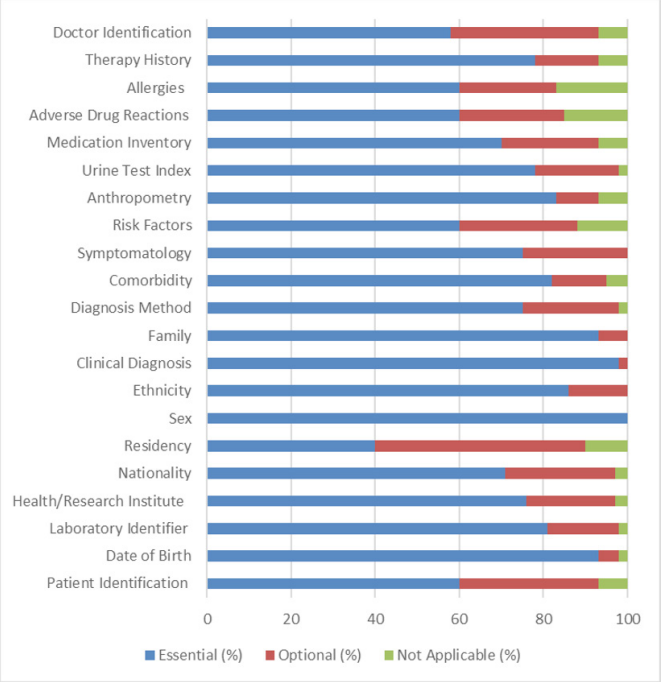

Figure 1 Survey response to proposed participant-level information.

and therapy. 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

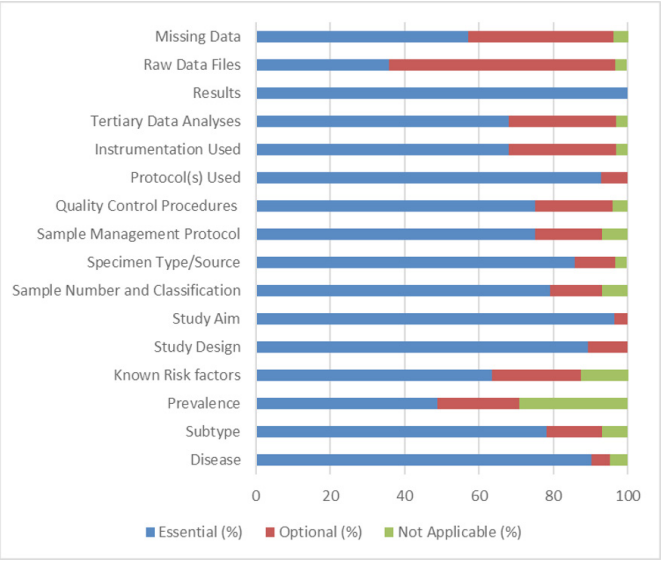

Figure 2 Survey response to proposed study- and experiment-level information.

Table 1 The Minimum Information Required Quideline: Kidney Disease Research and Clinical Data Reporting (version 1.0)

Participant-level information

Elements		Importance	Definition
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	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).
	Intensity		The average energy expended per physical activity. Light exercise is 20–60 min and elevates heart rate to 35%–60% of maximum heart rate (eg, housework, gardening, slow walking); moderate exercise is 20–60 min and elevates heart rate to 35%–60% of maximum heart rate (eg, basketball, single tennis, brisk walking); strenuous exercise elevates heart rate to over 60% of maximum heart rate (eg, 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).

Continued

Table 1 Continued

Participant-level information			
Elements		**Importance**	**Definition**
Anthropometrics	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 three measurements.
	Average weight	E	The measurement of mass or quantity of heaviness of an individual (in kg). Averaged over three 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 (ie, 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 mm Hg).
	Average diastolic blood pressure		The average pressure exerted into the systemic arterial circulation during cardiac ventricular relaxation and filling (in mm Hg).
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.
Urine and Serum Test Index	Urinary albumin	E	24 hours measurement of urine concentration of albumin (in mg/L).
	Urinary creatinine	E	24 hours measurement of urine concentration of creatinine (in mmol/L).
	Urinary total protein	E	24 hours measurement of urine concentration of total protein (in mg/L).
	Serum albumin	O	A quantitative spot-measurement of the amount of albumin present in a sample of serum (in µmol/L).
	Serum creatinine	O	A quantitative spot-measurement of the amount of creatinine present in a sample of serum (in µmol/L).
	Serum urea	O	A quantitative spot-measurement of the amount of creatinine present in a sample of serum. (in µmol/L).
	Estimated glomerular filtration rate	E	Measurement of the flow rate of filtered fluid through the kidney, calculated using the chronic kidney disease (CKD)-Epidemiology Collaboration (CKD-EPI) equation.
Kidney disease history	Personal history of kidney failure		
	Has a doctor or healthcare worker ever told you that you had kidney failure?	E	Participant's history of impaired kidney function, including kidney impairment, early-stage and advanced-stage kidney disease.
	How old were you when you were first told by a medical person that you had kidney failure?		
	If Yes, are one or both working well now?		
	Are you currently on renal dialysis?		
	Have you ever had a kidney transplant?		
	Family history of kidney failure		
	Has anyone in your family either had kidney disease or died from it?	E	Participant's background regarding impaired kidney function of blood relatives (persons related by descent).
	Do you know what type of kidney disease?		
	If Yes, please specify:'		
	Consanguinity		
	Any cases of consanguineous mating in the family?	O	Reproduction between genetically related individuals.
	'If Yes, please specify:'		

Continued

Table 1 Continued

Participant-level information			
Elements		**Importance**	**Definition**
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.
Kidney disease-related information	Hospital	E	The name and address of the institution that provides medical, surgical or psychiatric care and treatment for the sick or the injured
	Clinical diagnosis	E	The diagnosis made from the study of the signs and symptoms of a disease.
	Date of diagnosis	E	The calendar date on which a clinical diagnosis is made.
	Instrumentation	E	Specialised objects, or items of electrical or electronic equipment, employed to perform diagnosis (with versions).
	Clinical signs and symptoms	E	The objective evidence of disease perceptible to the examining healthcare worker (sign) and subjective evidence of disease perceived by the patient (symptom).
	Congenital conditions	O	The objective evidence of perceptible by the examining healthcare worker of defects that are present at birth.
	Histopathology	O	The visual examination of cells or tissue (or images of them) with an assessment regarding the quality of the cells or tissue.
	Comorbidities (systemic)	E	The presence of co-existing or additional systemic diseases with reference to an initial diagnosis or with reference to the index condition that is the subject of study.
	Comorbidities (pathogenic)	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.
Therapy	Therapy	E	Therapeutic intervention for kidney disease, excluding, drug therapy, diet therapy, radiotherapy and surgery.
	30-day frequency		The number of occurrences of kidney therapy per unit time (30 days).
	Start date		The calendar date on which treatment was initiated.
	End date		The calendar date on which treatment is to be or was terminated.
	End date		The calendar date on which treatment is to be or was terminated.
Study-level information			

Continued

Table 1 Continued

Participant-level information			
Elements		**Importance**	**Definition**
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.
	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).

E, essential; O, optional.

experimental aim, and optional elements such as output location which describes where the data will be saved.

The complete reporting guideline can be obtained from both the H3ABioNet website (www.h3abionet.org/data-standards/datastds) as well as FAIRsharing (https://fairsharing.org/bsg-s001385/), 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 while 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 utilisation and overlap with existing resources. A comprehensive guideline explaining how to employ the reporting guideline, along with the associated REDCap XML schema, locally can be found in online supplementary file 3. In addition, online supplementary file 4 contains an example entry of the reporting guideline, and online supplementary file 5 contains an illustration of the REDCap XML schema.

DISCUSSION

The Minimum Information Required Guideline: Kidney Disease Research and Clinical Data Reporting is a freely accessible, harmonised reporting guideline which can be employed or adapted for kidney disease research and healthcare and categorises information as essential or optional, as well as participant, study and experiment specific. Standardising how this information is captured, deposited, shared and published in a comparable and consistent manner is crucial for researchers to better understand a given study and subsequently interpret the data generated and conclusions made. The primary intent of the reporting guideline is to encourage harmonised data collection when launching new projects within the kidney disease research field. Ultimately, this will enhance the overall research community's capacity for conducting high-quality, interoperable and reusable research, adding long-term value to the collected clinical data and generated research data and encouraging more collaborative efforts worldwide. Similarly, the reporting guideline can also be employed retrospectively for data abstraction from existing or ongoing studies when reporting to a larger database, enabling the previously mentioned efforts.

Although certain elements within the standard can be incorporated into a case report form, the reporting guideline contains elements that need to be completed specifically by healthcare or research professionals, therefore the reporting standard is designed for use by research clinicians and healthcare workers, researchers, data managers and bioinformaticians involved in kidney disease research. The reporting guideline was not developed to replace the case report form but rather to provide a set of data reporting rules for researchers to adhere too. Defining the information as essential or optional permits the reporting guideline to be adaptable for both acquired and inherited kidney disease research, therefore elements such as congenital conditions and histopathology are defined as optional. The reporting standard goes beyond listing 'minimum required' data elements and aims to provide a comprehensive data dictionary, with standardised response options, which can be adapted for broad use. Therefore, employing the reporting standard allows comprehensive characterisation of research studies being conducted in the kidney disease research field, as well as the experiments and participants within these studies, supporting integrative analysis and improved biological interpretation.

The reporting standard is also accompanied by an associated REDCap XML schema. This was done to enable user friendliness and broad adoption of the standard as a data capturing and governance tool, allowing accurate and seamless duplication and reuse.[25] XML has been used extensively for describing data in many applications for storage or transport.[25] 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 is exchanged between healthcare systems.[25] It has previously been used in health reporting for such purposes.[26 27] Currently, ontologies cannot be intrinsically linked to the guideline elements within the REDCap XML. In the future, we aim to provide base XML schemas which are adaptable for broad implementation on various data capturing platforms. This will allow us to link the guideline elements to the mapped ontologies. Ultimately, the ontologies serve to promote FAIR reporting by adding an underlying layer of metadata and understanding to the overall dataset.

Broad adoption of the developed reporting standard has the potential to significantly reduce data and reporting inconsistency and redundancy across systems, promoting collaboration and(or) interoperability between projects.[28 29] Promoting such large-scale use could allow for improved data mapping in clinical registries, improving data quality and interoperability.[30] As previously exhibited in oncology research, broad adoption of a reporting standard can maximise the value and impact of research studies as well as the associated research data.[31] This is because research redundancy is reduced, and interpretable research outputs and comprehensive datasets are produced.[31] A given standard may be more widely adopted if advocated by databases, funding bodies and scientific journals, geared towards kidney disease research, specifically.

The Minimum Information Required Guideline: Kidney Disease Research and Clinical Data Reporting aims to promote FAIR reporting and will therefore be added to the FAIRsharing database, as this allows for continuous record maintenance and improvement, providing a point of contact for the standard, as well as related support material (https://doi.org/10.25504/FAIRsharing.fCAD2Z).[15] Bearing in mind the diverse target group the reporting standard aims to accommodate, various methods of implementation will be investigated to provide comprehensive

solutions for collaborative efforts. Additional elements will be investigated for incorporation into the standard, including environmental factors, dyslipidaemia and diet.

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. Ultimately, the reporting guideline has the potential to support both the H3Africa community as well as the kidney disease research community at large with current and future research.

Author affiliations
[1]H3ABioNet, Centre for Proteomic & Genomic Research, Cape Town, South Africa
[2]Department of Surgery, Division of Child Urology, University of Cape Town, Cape Town, South Africa
[3]Global Child Health Group, Amsterdam University Medical Centers, Amsterdam, The Netherlands
[4]KidGen Collaborative and AGHA Renal Genetics Flagships, Parkville, Victoria, Australia
[5]Institute of Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia
[6]Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia
[7]Department of Parasitology, Institut Pasteur de Tunis, Tunis, Tunisia
[8]National Institute of Mathematical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[9]Department of Digital Technologies, University of Mauritius, Reduit, Mauritius
[10]Department of Integrative Biomedical Sciences, Division of Computational Biology, University of Cape Town, Cape Town, South Africa

Twitter Zahra Mungloo-Dilmohamud @ZahraMungloo

Acknowledgements The authors would like to acknowledge the kidney disease researchers, clinicians and hospital departments who participated and contributed to the consolidation and finalisation of the reporting guideline. Patients were not involved in the reported project.

Contributors The current project was designed by JK and supervised by NM. LZ, MC, CVW and AM were the primary developers of the reporting guidelines and drafted the manuscript. KJ, SB, ZMD and CP drafted and corrected the manuscript. MT and MM were the primary developers of the XML files. All authors read and approved the final manuscript.

Funding H3ABioNet is supported by the National Institutes of Health Common Fund under grant number U41HG006941. 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. The Australian Genomics Health Alliance is supported by the National Health and Medical Research Council under grant number APP1113531.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

REFERENCES

1 Fan J, Han F, Liu H. Challenges of big data analysis. *Natl Sci Rev* 2014;1:293–314.
2 Luo J, Wu M, Gopukumar D, *et al*. Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights* 2016;8:BII.S31559–10.
3 Lee CH, Yoon H-J. Medical big data: promise and challenges. *Kidney Res Clin Pract* 2017;36:3–11.
4 Floege J, Mak RH, Molitoris BA, *et al*. Nephrology research--the past, present and future. *Nat Rev Nephrol* 2015;11:677–87.
5 Hildebrandt F, diseases Gkidney. Genetic kidney diseases. *The Lancet* 2010;375:1287–95.
6 Köttgen A, Pattaro C, Böger CA, *et al*. New loci associated with kidney function and chronic kidney disease. *Nat Genet* 2010;42:376–84.
7 Mallett AJ, McCarthy HJ, Ho G, *et al*. Massively parallel sequencing and targeted exomes in familial kidney disease can diagnose underlying genetic disorders. *Kidney Int* 2017;92:1493–506.
8 Anand S, Stanifer JW, Thomas B. Clinical nephrology research in low-resource settings: opportunities, priorities, and challenges for young Investigators. *Clin Nephrol* 2016;86:8–13.
9 Bello AK, Levin A, Tonelli M, *et al*. Assessment of global kidney health care status. *JAMA* 2017;317:1864–81.
10 Chervitz SA, Deutsch EW, Field D, *et al*. Data standards for omics data: the basis of data sharing and reuse. *Methods Mol Biol* 2011;719:31–69.
11 Khan AM. Guidelines for standardizing and increasing the transparency in the reporting of biomedical research. *J Thorac Dis* 2017;9:2697–702.
12 Skrocki M. Standardization needs for effective Interoperability. *Trans Int Conf Health Inf Technol Adv* 2013.
13 Kim K. Clinical Data Standards in Health Care: Five Case Studies. In: *California health care Foundation*, 2005. http://www.chcf.org/
14 Hamilton CM, Strader LC, Pratt JG, *et al*. The PhenX toolkit: get the most from your measures. *Am J Epidemiol* 2011;174:253–60.
15 Sansone S-A, McQuilton P, Rocca-Serra P, *et al*. FAIRsharing as a community approach to Standards, repositories and policies. *Nat Biotechnol* 2019;37:358–67.
16 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3. 201618.
17 Cole NI, Liyanage H, Suckling RJ, *et al*. An ontological approach to identifying cases of chronic kidney disease from routine primary care data: a cross-sectional study. *BMC Nephrol* 2018;19:85.
18 Köhler S, Vasilevsky NA, Engelstad M, *et al*. The human phenotype ontology in 2017. *Nucleic Acids Res* 2017;45:D865–76.
19 Stearns MQ, Price C, Spackman KA, *et al*. Snomed clinical terms: overview of the development process and project status. *Proc AMIA Symp* 2001:662–6.
20 Mulder NJ, Adebiyi E, Alami R, *et al*. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Res* 2016;26:271–7.
21 Rotimi C, Abayomi A, Abimiku Alash'le, *et al*. Research capacity. enabling the genomic revolution in Africa. *Science* 2014;344:1346–8.
22 Noy NF, Shah NH, Whetzel PL, *et al*. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009;37:W170–3.
23 Côté RG, Jones P, Apweiler R, *et al*. The ontology Lookup service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* 2006;7:97.
24 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
25 Eito-Brun R. XML: The Basis of the Language. In: Eito-Brun R, ed. *XML-based content management*. Chandos Publishing, 2018: 1–30.
26 Huser V, Sastry C, Breymaier M, *et al*. 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* 2015;57:88–99.
27 Schweiger R, Brumhard M, Hoelzer S, *et al*. Implementing health care systems using XML standards. *Int J Med Inform* 2005;74:267–77.
28 Hartwell SW, Deng X, Fisher W, *et al*. Harmonizing databases? developing a quasi-experimental design to evaluate a public mental health re-entry program. *Eval Program Plann* 2012;35:461–72.
29 Pool R, Esnayra J, Bioinformatics: Converting Data to Knowledge: Workshop Summary. *Barriers to the use of databases*. US: National Academies Press, 2000.
30 Rastegar-Mojarad M, Sohn S, Wang L, *et al*. Need of informatics in designing interoperable clinical registries. *Int J Med Inform* 2017;108:78–84.
31 MacCarthy A, Kirtley S, de Beyer JA, *et al*. Reporting guidelines for oncology research: helping to maximise the impact of your research. *Br J Cancer* 2018;118:619–28.